

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds
(without alignments)
1433.395 Million cell updates/sec

Title: US-09-887-784-64A
Perfect score: 1273
Sequence: 1 MVSGEELFTGVVPILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04: *
1: geneseqp1980s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003bs: *
8: geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1269	99.7	239	5	Aae17518 Enhanced
2	1269	99.7	363	6	ABR40352 Human aml
3	1269	99.7	893	4	AAG65781 Amino aci
4	1269	99.7	1132	4	AAG65782 Amino aci
5	1261	99.1	239	5	Aae17517 Enhanced
6	1258	98.8	239	3	AAB22882 Enhanced
7	1258	98.8	239	3	AAY54349 Amino aci
8	1258	98.8	239	3	AAY79584 EGFP sign
9	1258	98.8	239	4	AAB50804 Jellyfish
10	1258	98.8	239	4	AAB85900 A. victor
11	1258	98.8	239	4	AAB31171 Amino aci
12	1258	98.8	239	5	AAG66198 A. victor
13	1258	98.8	239	5	ABG94444 Protease
14	1258	98.8	239	5	Aae14599 Aequorea
15	1258	98.8	239	6	Aae34958 Aequorea
16	1258	98.8	239	6	AAG79829 Green flu
17	1258	98.8	239	6	ABR83616 Green flu
18	1258	98.8	239	6	ADA38074 Aequorea
19	1258	98.8	239	7	ABU63204 Aequorea
20	1258	98.8	239	7	ADC18358 EGFP (enh
21	1258	98.8	239	7	ABW00914 Aequorea
22	1258	98.8	239	7	ADE28570 Enhanced
23	1258	98.8	246	7	ABM79011 Enhanced
24	1258	98.8	248	5	AAG68319 Jellyfish
25	1258	98.8	259	5	Aau99804 Biomembra

26	1258	98.8	265	2	AAW97451 Wild-type
27	1258	98.8	268	5	AAU99803 Biomembra
28	1258	98.8	270	5	AAU99802 Biomembra
29	1258	98.8	272	5	AAU99800 Biomembra
30	1258	98.8	273	5	AAU99801 Biomembra
31	1258	98.8	280	5	AAU99807 Biomembra
32	1258	98.8	281	3	AAU50142 Green flu
33	1258	98.8	281	3	AAE24252 EGFP-MODC
34	1258	98.8	281	5	AAU10888 EGFP-MODC
35	1258	98.8	286	7	ADE28562 EGFP/ hum
36	1258	98.8	289	7	ADE28564 EGFP/ hum
37	1258	98.8	290	7	ADE28568 EGFP/ hum
38	1258	98.8	290	7	ADE28566 EGFP/ hum
39	1258	98.8	294	3	AAE22860 GFP-DEVD-
40	1258	98.8	294	3	AAU79638 Caspase-3
41	1258	98.8	294	5	ABG94422 Recombina
42	1258	98.8	308	2	AAU42181 EGFP/DRM
43	1258	98.8	320	6	ABR83620 HUB1-GFP
44	1258	98.8	323	3	AAU54359 GFP mutan
45	1258	98.8	323	6	ABR83621 HUB1-GFP

ALIGNMENTS

RESULT 1
AAE17518
ID AAE17518 standard; protein; 239 AA.
AC AAE17518;
XX
XX
DT 22-APR-2002 (first entry)
XX
DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.
XX
KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
KW cellular function; genetic reporter; mutant; Stoke's shift; muten.
XX
OS Aequorea victoria.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 65 /note= "Wild type Phe substituted with Leu; This
FT corresponds to position 64 in the wild type protein"
FT Misc-difference 223
FT /note= "Wild type Glu substituted with Gly; This
FT corresponds to position 222 in the wild type protein"

WO200198338-A2.
27-DEC-2001.
18-JUN-2001; 2001WO-EP006848.
19-JUN-2000; 2000DK-00000953.
20-JUN-2000; 2000US-0212681P.
10-MAY-2001; 2001DK-00000739.
10-MAY-2001; 2001US-0290170P.
(BIOI-) BIOIMAGE AS.
Bjorn SP, Pagliaro L, Thastrup O;
WPI; 2002-098224/13.
N-PSDB; AAD28163.
Novel fluorescent protein in vitro assay for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution, has a green fluorescent protein with F64L and E222G mutation.
Claim 9; Page 37; 41pp; English.

XX The invention relates to a fluorescent protein derived from green
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
CC at F64L and E222G has a bigger compared to other GFP's making it very
CC suitable for high throughput screening due to better resolution. The
CC fluorescent protein is useful in *in vitro* assays for measuring protein
CC kinase activity or dephosphorylation activity, or for measuring protein
CC redistribution. The fluorescent protein is useful in studying cellular
CC functions in living cells; as protein tags in transgenic animals, living
CC and fixed cells; organelle tags, secretion marker and genetic reporter.
CC The fluorescent protein is also useful as a cell or organelle integrity
CC marker, a marker for changes in cell morphology, as transfection marker,
CC and as a marker to be used in combination with fluorescence activated
CC cell sorting (FACS). The novel proteins can also be used as reporters to
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
CC protein is also useful as markers in transcriptional and translational
CC fusions for performing transposon vector mutagenesis and as a reporter
CC for bacterial detection. Transposons encoding the fluorescent protein are
CC useful for screening promoters and for tagging plasmids and chromosomes.
CC The fluorescent protein engineered into the genome of a phage is useful
CC for designing diagnostic tool. The present sequence is a DNA encoding
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
XX
XX Sequence 239 AA;

Query Match 99.7%; Score 1269; DB 5; Length 239;
Best Local Similarity 99.6%; Pred. No. 4.3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEBELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEBELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFRSYRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFRSYRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239

RESULT 2
ABR40352
ID ABR40352 standard; protein; 363 AA.

XX ABR40352;
AC ABR40352;
XX 08-JUL-2003 (first entry)
DT
DE Human amino acid sequence SEQ ID NO: 6.
XX Human; heterologous conjugate; intracellular protein.
XX Homo sapiens.
XX Aequorea victoria.
XX WO2003029827-A2.
XX 10-APR-2003.
XX 01-OCT-2002; 2002WO-DK000651.
XX 01-OCT-2001; 2001DK-00001433.
XX 11-OCT-2001; 2001US-0328896P.
XX (BIOL-) BIOLMAGE AS.
XX Terry ER, Nielsen SJ;

XX WPI; 2003-430211/40.
DR N-PSDB; ACC72604.

XX Novel cell for identifying modulators of protein interaction, contains a
PT first conjugate comprising anchor protein, second conjugate having type B
PT interactor protein and third conjugate with detectable group.
XX
XX Disclosure; Page 112-113; 118pp; English.

XX The invention relates to a novel cell, comprising three heterologous
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that
CC specifically binds to an internal structure within the cell conjugated to
CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
CC type B conjugated to a first protein of interest, and a third HC (HC3)
CC comprising a second protein of interest conjugated to detectable group.
CC The cell is useful for detecting if a compound disrupts or induces the
CC interaction between two intracellular proteins. The cell is also useful
CC for screening compounds that modulate the interaction between two
CC intracellular proteins. The present sequence is used in the
CC exemplification of the invention
XX
XX Sequence 363 AA;

Query Match 99.7%; Score 1269; DB 6; Length 363;
Best Local Similarity 99.6%; Pred. No. 8.1e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEBELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEBELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFRSYRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFRSYRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239

RESULT 3
AAG65781
ID AAG65781 standard; protein; 893 AA.

XX AAG65781;
AC AAG65781;
XX 07-JAN-2002 (first entry)
DT
DE Amino acid sequence of HSPDE4A1-E222G fusion protein.
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX fusion protein.
XX Homo sapiens.
XX Aequorea victoria.
XX WO200179526-A2.
XX 25-OCT-2001.
XX 11-APR-2001; 2001WO-DK000264.
XX 17-APR-2000; 2000DK-00000651.
XX 29-MAY-2000; 2000DK-00000849.
XX (BIOI-) BIOIMAGE AS.


```
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66852.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 156-160; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as, depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
XX protein
XX
XX Sequence 893 AA;
XX
XX Query Match 99.7%; Score 1269; DB 4; Length 893;
XX Best Local Similarity 99.6%; Pred. No. 3.1e-123;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
Db 655 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 714
XX
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 715 LVTALSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
XX
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
XX
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 893
XX
RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA;
XX
AC AAG65782;
XX
XX 07-JAN-2002 (first entry)
XX
DE Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW fusion protein.
XX
XX Homo sapiens.
OS Aequorea victoria.
XX
XX WO200179526-A2.
PN
XX
```

```
PD 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-DK000264.
XX
XX 17-APR-2000; 2000DK-00000651.
XX 29-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
XX Praestegaard M;
PI
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66853.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 162-167; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as, depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
XX protein
XX
XX Sequence 1132 AA;
XX
XX Query Match 99.7%; Score 1269; DB 4; Length 1132;
XX Best Local Similarity 99.6%; Pred. No. 4.5e-123;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
Db 894 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 953
XX
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 954 LVTALSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013
XX
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
XX
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 1132
XX
RESULT 5
AAE17517
ID AAE17517 standard; protein; 239 AA.
XX
XX AAE17517;
XX
XX 22-APR-2002 (first entry)
XX
XX Enhanced F64L jellyfish green fluorescent protein mutant.
XX
```


CC biosensors of the invention can be used to investigate a wide range of
CC cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased quantity
CC and speed of data collection and faster evaluation of drug candidates.
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used
CC as components of biosensor fusion proteins of the invention
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYVKLTIFCTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYVKLTIFCTTGKLPVPWPT 60

QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 7
AAY54349
ID AAY54349 standard; protein; 239 AA.
XX
AC AAY54349;
XX
DT 06-APR-2000 (first entry)
XX
DE Amino acid sequence of the mutant green fluorescent protein EGFP.
XX
KW Fluorescent protein; green fluorescent protein; emission intensity;
KW fluorescence; pH detection; pH sensor; EGFP.
XX
OS Synthetic.
OS Aequorea victoria.
XX
FH Key Location/Qualifiers
FT Misc-difference 65 /note= "wild type Phe substituted with Leu"
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"
FT Misc-difference 232 /note= "wild type His substituted with Leu"
XX
XX
PN WO9964592-A2.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US012850.
XX
XX 09-JUN-1998; 98US-00094359.
PR 13-OCT-1998; 98US-00172063.
XX
XX (REGC) UNIV CALIFORNIA.
PA (UYOR-) UNIV OREGON STATE.
XX

PI Tsien RY, Llopis J, Wachter RM;
XX WPI; 2000-116540/10.
DR N-PSDB; AA245642.
XX
PT New functional engineered green fluorescent proteins, used for measuring
PT the pH in biological samples and cells.
XX
PS Disclosure; Page 9; 89pp; English.
XX
CC The present sequence represents a functional engineered fluorescent
CC protein based on the Aequorea green fluorescent protein (GFP). The
CC emission intensity changes as pH varies between 5 and 10 of the present
CC protein are novel. The functional engineered fluorescent proteins show
CC reversible changes in fluorescence over physiological pH ranges. They can
CC be used for determining the pH of samples and cells. The polynucleotides
CC can also be used to produce transgenic animals. The fluorescent protein
CC pH sensors can be delivered to cells in the form of polynucleotides
CC encoding the protein sensor fused to a targeting signal. The targeting
CC signal directs the expression of the protein sensors to restricted cell
CC locations. This makes it possible to measure the pH of a precisely
CC defined cellular region or organelle
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYVKLTIFCTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYVKLTIFCTTGKLPVPWPT 60

QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 8
AAY79584
ID AAY79584 standard; peptide; 239 AA.
XX
AC AAY79584;
XX
DT 29-AUG-2000 (first entry)
XX
DE EGFP signal domain.
XX
KW Protease; biosensor; EGFP; signal peptide; cell screening; assay;
KW analysis; drug discovery.
XX
OS Unidentified.
XX
PN WO200026408-A2.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99WO-US025431.
XX
XX 30-OCT-1998; 98US-0106308P.
PR 26-MAY-1999; 99US-0136078P.
XX
XX (CELL-) CELLOMICS INC.
XX
PI Giuliano KA, Bright G, Olson K, Burroughs-Tencza S;

WPI; 2000-365644/31.
N-PSDB; AAA27573.

Recombinant nucleic acid encoding a protease biosensor useful for fluorescence based cell and molecular biochemical assays for drug discovery comprising three operably linked nucleic acid sequences.

Claim 14; Fig 29A; 218pp; English.

The present sequence is that of the EGFP signal domain, which can be included in novel recombinant protease biosensors (PBs) of the invention. The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87) comprising at least 1 detectable polypeptide signal such as the present sequence; a second domain (see AAY79586-622) comprising at least 1 protease recognition site; and a third domain (see AAY79623-37) comprising at least 1 reactant target sequence. A recombinant nucleic acid (see AAA7627-43) encoding the PB, an expression vector, and a genetically engineered host cell are also claimed. A claimed method for identifying compounds that modify protease activity in a cell involves contacting a host cell that possesses the recombinant PB with a test compound, and determining the PB distribution in the host cell, where changes in the distribution of the PB are correlated with modification of protease activity by the test compound. Claimed kits for identifying compounds that modify protease activity in a host cell include the recombinant nucleic acid, or the recombinant PB, or the vector, or the host cell. The PB is useful in high content screens to detect *in vivo* activation of enzymatic activity, and to identify specific activity based on cleavage of a known recognition motif

Sequence 239 AA;
SQ

	Query Match	98.8%;	Score 1258;	DB 3;	Length 239;
	Best Local Similarity	98.7%;	Pred. No. 6.1e-133;		
	Matches 236;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	MVSGBELFTGVVPI	LVLELGDVNGHKFVS	GEGEDATY	GKLTLCICTTGKLPVPWPT 60
DB	1	MVSGBELFTGVVPI	LVLELGDVNGHKFVS	GEGEDATY	GKLTLCICTTGKLPVPWPT 60
QY	61	LVTALSYGVQCF	SRYPDHMKQHD	FFKSAMPEGVY	QVQERTIIPFKDDGNKYTRAEVFE
DB	61	LVTTLTYGVQCF	SRYPDHMKQHD	FFKSAMPEGVY	QVQERTIIPFKDDGNKYTRAEVFE
QY	121	VNRIELKGI	DFKEDGNLTGHKLE	VYNNYSHNVY	INADKQKNGIKVNFKIRHNI
DB	121	VNRIELKGI	DFKEDGNLTGHKLE	VYNNYSHNVY	INADKQKNGIKVNFKIRHNI
QY	181	DHYQONTPI	IGDGPVILLDP	NHLYSTQSALS	KDNPNEKRDHMLLVFTAA
DB	181	DHYQONTPI	IGDGPVILLDP	NHLYSTQSALS	KDNPNEKRDHMLLVFTAA

RESULT 9	
AAB50804	
ID	AAB50804 standard; protein; 239 AA.
XX	
XX	
XX	AAB50804;
XX	AC
XX	14-MAR-2001 (first entry)
DT	
DT	Jellyfish GFP mutant EGFP.
XX	
DE	
XX	
XX	Aequorea victoria; jellyfish; fluorescent protein indicator;
XX	green fluorescent protein; GFP; linker moiety; sensor;
KW	calmodulin-binding domain; mutant; mutein.
KW	
XX	
XX	Aequorea victoria.
OS	
XX	
PN	WO200071565-A2.
XX	
XX	
PD	30-NOV-2000.
XX	
XX	

17-MAY-2000; 2000WO-US013684.
21-MAY-1999; 99US-00316919.
21-MAY-1999; 99US-00316920.
(REGC) UNIV CALIFORNIA.
Tsien RY, Baird GA;
WPI; 2001-032017/04.
N-PSDB; AAC90488.
Novel fluorescent proteins comprising a sensor protein inserted into
them, useful for measuring the response of a sensor biological, chemical,
electrical or physiological parameter in vivo or in vitro.
Disclosure; Page 24; 94pp; English.
The present sequence is a fluorescnat protein used in the construction of
a fluorescent protein indicator. The indicator comprises a sensor
polypeptide that is responsive to a chemical, biological, electrical or
physiological parameter, and a fluorescence protein functional group. The
sensor polypeptide is operatively inserted into the fluorescent moiety.
The fluorescent indicator is useful for detecting the presence of a
response inducing member in a sample. The method involves contacting the
sample with the indicator and detecting a change in fluorescence, in
which a change is indicative of the effect of the parameter on the sensor
polypeptide. The novel fluorescent proteins are advantageous due to their
reduced size as compared to the FRET (fluorescence resonance energy
transfer)-based sensors
Sequence 239 AA;

Query Match	98.8%;	Score 1258;	DB 4;	Length 239;	
Best Local Similarity	98.7%;	Pred. No. 6.1e-123;			
Matches 236;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0	
Qy	1	MVSGEELFTGVVPI	LVELDGVNGHFKFSVSGEGGDATY	GKLTILKFKICTTGKLPVWPPT 60	
Db	1	MVSGEELFTGVVPI	LVELDGVNGHFKFSVSGEGGDATY	GKLTILKFKICTTGKLPVWPPT 60	
Qy	61	LVLTALSYGVQCF	SPYDPHMKQHDFFK	SAMPYGYVQERTIFFKDDGNYKTRAEVKEGDTL 120	
Db	61	LVTTLTYGVCFS	PYDPHMKQHDFFK	SAMPYGYVQERTIFFKDDGNYKTRAEVKEGDTL 120	
Qy	121	VNRIELKGI	DFKEDCNILGHKLEYNYN	SHVYIIMADKQKNGIKVNFKIRHNI	EDGSVQLA 180
Db	121	VNRIELKGI	DFKEDCNILGHKLEYNYN	SHVYIIMADKQKNGIKVNFKIRHNI	EDGSVQLA 180
Qy	181	DHYQONTPI	GDGPVLLPDN	HYLSTQSALSKDPN	KEKRDHMLLGFVTAAGITLGMDELYK 239
Db	181	DHYQONTPI	GDGPVLLPDN	HYLSTQSALSKDPN	KEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 10	
AAH85900	
ID	AAH85900 standard; protein; 239 AA.
XX	
XX	AAH85900;
XX	
XX	30-NOV-2001 (first entry)
DT	
DT	
XX	
XX	A. victoria green fluorescent protein (GFP) and linker sequence.
DE	
XX	
XX	Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
KW	fluorescent polypeptide; orexinergic; anabolic; food intake; GFP;
KW	green fluorescent protein.
XX	
XX	
OS	Synthetic.
OS	Aequorea victoria.
XX	
XX	WC0200168706-A1.
PN	

```
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-US008071.
XX
PR 15-MAR-2000; 2000US-0189698P.
XX
PA (MERI ) MERCK & CO INC.
XX
PI Marsh DJ;
XX
XX WPI; 2001-565791/63.
XX
DR N-PSDB; AAH47304.
XX
XX Fusion proteins comprising melanin concentrating hormone receptor
PT peptides and fluorescent proteins, useful for identifying appetite
PT stimulants.
XX
PS Claim 2; Page 14; 71pp; English.
XX
CC The invention provides melanin concentrating hormone (MCH) receptor
CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
CC MCHR polypeptide regions from different species. The MCHR fusion protein
CC comprise MCHR polypeptide region and a fluorescent polypeptide region
CC joined directly, or via a linker, to the carboxy side of the MCHR
CC polypeptide region. The MCHR fusion proteins can be expressed by standard
CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
CC regulation of MCH activity stimulates food intake. The present sequence
CC represents a A. victoria green fluorescent protein (GFP) and a linker
CC sequence
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 11
AAB311171
ID AAB311171 standard; protein; 239 AA.
XX
AC AAB311171;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of a green fluorescent protein (GFP).
XX
KW Growth rate; death rate; reporter gene; luminescent protein;
KW fluorescent product; luciferase; green fluorescent protein; GFP.
XX
OS Aequorea victoria.
XX
PN WO200075367-A1.
XX
PD 14-DEC-2000.
XX
PF 07-JUN-2000; 2000WO-FI000507.

XX 07-JUN-1999; 99FI-00001296.
XX
XX (LILI/) LILIUS E.
XX
XX (VIRT/) VIRT M.
XX
XX Lilius E, Virta M;
XX
XX WPI; 2001-061737/07.
XX
XX N-PSDB; AAC86954.
XX
XX Assessing growth and death rates of a micro-organism in a desired
XX environment, by introducing 2 reporter genes encoding luminescent and
XX fluorescent products and detecting luminescent fluorescence.
XX
XX Disclosure; Page 27; 32pp; English.
XX
XX The specification describes a method for assessing the growth rate and
XX death rate of a micro-organism within a predetermined time period in a
XX desired environment. The method comprises introducing at least two
XX reporter genes encoding luminescent and/or fluorescent products into the
XX micro-organisms, incubating the micro-organism within the desired
XX environment, and detecting luminescence and/or fluorescence after a
XX predetermined time period. Use of two different markers within a micro-
XX organism enables the differentiation between growth and death rates. The
XX method is used to assess the growth rate and death rate of a micro-
XX organism within a predetermined time period in a desired environment. The
XX present sequence represents a green fluorescent protein (GFP), and is
XX encoded by a plasmid which encodes luminescent and fluorescent proteins,
XX and is used in the method of the invention
XX
XX Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 12
AAG66198
ID AAG66198 standard; protein; 239 AA.
XX
AC AAG66198;
XX
DT 17-JUN-2002 (first entry)
XX
DE A. victoria green fluorescent protein (EGFP).
XX
KW Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
KW green fluorescent protein; EGFP.
XX
OS Aequorea victoria.
XX
PN JP2002045189-A.
XX
PD 12-FEB-2002.
XX
```

```

PF 04-AUG-2000; 2000JP-00237165.
XX
PR 04-AUG-2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2002-299190/34.
XX N-PSDB; ABL40628.
XX
XX A gene encoding cyan-green fluorescent protein.
XX
XX Example; Page 14; 20pp; Japanese.
XX
XX The invention relates to a gene encoding proteins having cyan-green
XX fluorescence characteristic and having a function of showing stable
XX fluorescence characteristic in acid region. A method for the preparation
XX of a cyan-green fluorescent protein is provided which involves a
XX transforment transformed by a recombinant vector comprising the gene,
XX where the transformant is cultured and the protein is collected from the
XX culture. The present sequence represents the A. victoria green
XX fluorescent protein (EGFP)
XX
XX Sequence 239 AA;
SQ
Query Match 98.8%; Score 1258; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MWSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVWPWT 60
DB 1 MWSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVWPWT 60
QY 51 LVTALSYGQCFSRYPDHMKQHDFFKSPAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSPAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQNTPTIGDGPVLLPDNHYLSTQSALSCKDNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQNTPTIGDGPVLLPDNHYLSTQSALSCKDNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 13
ABG94444
ID ABG94444 standard; protein; 239 AA.
XX
XX ABG94444;
XX
XX 27-NOV-2002 (first entry)
XX
XX Protease biosensor signal sequence #6.
XX
XX Detection; classification; identification; toxin detection; protease;
XX ADP-ribosylating toxin; cytotoxic phospholipase; excitatory toxin;
XX toxic threat agent.
XX
XX Synthetic.
XX
XX US6416959-B1.
XX
XX 09-JUL-2002.
XX
XX 25-FEB-2000; 2000US-00513783.
XX
XX 27-FEB-1997; 97US-00810983.
XX 27-FEB-1998; 98US-00031271.
XX 26-FEB-1999; 99US-0122152P.
XX 08-MAR-1999; 99US-0123399P.
XX 12-JUL-1999; 99US-00352171.
XX 31-AUG-1999; 99US-0151797P.
XX

```

```

PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
PR 01-DEC-1999; 99US-0168408P.
XX
XX (GIUL/) GIULIANO K.
XX (KAPU/) KAPUR R.
XX
XX Giuliano K, Kapur R;
XX
XX WPI; 2002-634730/68.
XX N-PSDB; ABS71491.
XX
XX Automated cell-based toxin detection, classification, and/or
XX identification by treating cells involves use of three classes of
XX luminescent reporter molecules such as detectors, classifiers or
XX identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
XX
XX The invention describes methods of automated detection, classification
XX and identification comprising treating cells containing luminescent
XX reporter molecules (1) in array of locations with a test substance, where
XX (1) are detectors, classifiers or identifiers, imaging cells in each
XX location to obtain luminescent signals and converting optical information
XX into digital data to interpret presence of toxins in the test substance.
XX The method are useful for detection of toxins chosen from proteases, ADP-
XX ribosylating toxins, cytotoxic phospholipases, and excitatory toxins.
XX Three classes of cell-based luminescent reporter molecules such as
XX detectors, classifiers and identifiers are described and serve as
XX reporters of toxic threat agents. The first two levels of
XX characterisation enable a rapid readout of toxin class without
XX sacrificing the ability to detect many new mutant toxins or dissect
XX several complex mixtures of known toxins. This is the amino acid sequence
XX of a protease biosensor related signal sequence used in the cell-based
XX screening system
XX
XX Sequence 239 AA;
SQ
Query Match 98.8%; Score 1258; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MWSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVWPWT 60
DB 1 MWSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVWPWT 60
QY 61 LVTALSYGQCFSRYPDHMKQHDFFKSPAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSPAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSCKDNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSCKDNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AAE14599;
XX
XX 31-MAY-2002 (first entry)
XX
XX Aequorea victoria enhanced green fluorescent protein.
XX
XX Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; muten.
XX
XX Aequorea victoria.
XX
XX Synthetic.

```

XX Key Location/Qualifiers
 FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"
 FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"
 FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"
 XX EP1178109-A1.
 XX 06-FEB-2002.
 XX 03-AUG-2001; 2001EP-00306650.
 XX 04-AUG-2000; 2000JP-00237166.
 XX (RIKE) RIKEN KK.
 XX Miyawaki A, Sawano A;
 XX WPI; 2002-208112/27.
 XX N-PSDB; AAD27910.
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.
 XX Example 1; Page 13-14; 31pp; English.
 XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria
 XX Sequence 239 AA;
 XX Query Match 98.8%; Score 1258; DB 5; Length 239;
 XX Best Local Similarity 98.7%; Pred. No. 6.1e-123;
 XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQQNTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

QY 181 DHYQQNTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
 RESULT 15
 AAE34958
 ID AAE34958 standard; protein; 239 AA.
 XX AAE34958;
 AC AAE34958;
 XX 28-MAY-2003 (first entry)
 XX Aequorea victoria enhanced green fluorescent protein (EGFP).
 DE Phosphorylation indicator; fluorescent protein; detection; phosphatase;
 KW Kinase; enhanced green fluorescent protein; EGFP.
 OS Aequorea victoria.
 XX WO200295058-A2.
 XX 28-NOV-2002.
 XX 24-MAY-2002; 2002WO-USO16955.
 XX 24-MAY-2001; 2001US-00865291.
 XX (REGC) UNIV CALIFORNIA.
 XX Tsien RY, Ting AY, Zhang J;
 XX WPI; 2003-148474/14.
 XX N-PSDB; AAD53428.
 XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphaminoacid binding domain, and acceptor molecule, in operative linkage.
 XX Disclosure; Col 56-57; 38pp; English.
 XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or operative linkage, a donor molecule, a phosphorylatable domain, a phosphaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention
 XX Sequence 239 AA;
 XX Query Match 98.8%; Score 1258; DB 6; Length 239;
 XX Best Local Similarity 98.7%; Pred. No. 6.1e-123;
 XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQQNTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

us-09-887-784-64a.rag

Sun Jun 27 18:27:35 2004

Db 181 DHYQONTPIGDGEVLLPDNHYLSTQSALSKDPNEXKRDHNVLLFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:09
Job time : 48.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-64A

Perfect score: 1273
Sequence: 1 MYSKGEELFTGVVPIVLVD.....VLLGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1258	98.8	239	3	US-09-172-063-3
2	1258	98.8	239	4	US-09-513-783A-46
3	1258	98.8	239	4	US-09-316-919-4
4	1258	98.8	239	4	US-09-602-641-3
5	1258	98.8	239	4	US-09-920-922-2
6	1258	98.8	281	3	US-09-062-102-1
7	1258	98.8	281	4	US-09-364-946-1
8	1258	98.8	294	4	US-09-513-783A-2
9	1258	98.8	323	3	US-09-172-063-21
10	1258	98.8	323	4	US-09-602-641-21
11	1258	98.8	364	3	US-09-085-305-6
12	1258	98.8	379	4	US-09-417-197-129
13	1258	98.8	434	4	US-09-800-170-48
14	1258	98.8	442	4	US-09-417-197-127
15	1258	98.8	459	4	US-09-513-783A-170
16	1258	98.8	544	4	US-09-417-197-113
17	1258	98.8	544	4	US-09-417-197-115
18	1258	98.8	604	4	US-09-417-197-59
19	1258	98.8	605	4	US-09-417-197-41
20	1258	98.8	606	4	US-09-417-197-65
21	1258	98.8	607	4	US-09-417-197-47
22	1258	98.8	630	4	US-09-417-197-63
23	1258	98.8	631	4	US-09-417-197-39
24	1258	98.8	633	4	US-09-417-197-45
25	1258	98.8	635	4	US-09-417-197-125
26	1258	98.8	642	2	US-08-818-253-2
27	1258	98.8	642	2	US-08-818-253-6

RESULT 1
US-09-172-063-3
; Sequence 3, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Liopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-172-063-3

Query Match 98.8%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MYSKGEELFTGVVPIVLVDGVNKHFSVSGEGDATYVKLTLPKFTCTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPIVLVDGVNKHFSVSGEGDATYVKLTLPKFTCTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNVKTAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNVKTAEVKFEGDTL 120
QY 121 VNRIELKIDFKEDGNILGHKLEYNVNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180
DB 121 VNRIELKIDFKEDGNILGHKLEYNVNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180
QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

ALIGNMENTS

28 1258 98.8 642 3 US-08-818-252-2 Sequence 2, Appli
29 1258 98.8 642 3 US-08-818-252-6 Sequence 6, Appli
30 1258 98.8 652 2 US-08-818-253-4 Sequence 4, Appli
31 1258 98.8 652 3 US-08-818-253-4 Sequence 4, Appli
32 1258 98.8 718 4 US-09-417-197-75 Sequence 75, Appli
33 1258 98.8 719 4 US-09-417-197-51 Sequence 51, Appli
34 1258 98.8 726 4 US-09-417-197-71 Sequence 71, Appli
35 1258 98.8 727 4 US-09-417-197-139 Sequence 139, App
36 1258 98.8 783 4 US-09-513-783A-176 Sequence 176, App
37 1258 98.8 797 4 US-09-417-197-141 Sequence 141, App
38 1258 98.8 797 4 US-09-417-197-143 Sequence 143, App
39 1258 98.8 798 4 US-09-417-197-77 Sequence 77, Appli
40 1258 98.8 805 4 US-09-513-783A-178 Sequence 178, App
41 1258 98.8 806 4 US-09-417-197-53 Sequence 53, Appli
42 1258 98.8 836 4 US-09-417-197-61 Sequence 61, Appli
43 1258 98.8 842 4 US-09-417-197-43 Sequence 43, Appli
44 1258 98.8 843 4 US-09-417-197-117 Sequence 117, App
45 1258 98.8 853 4 US-09-417-197-119 Sequence 119, App

```
RESULT 2
US-09-513-733A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-733A-46

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGGKLPVPWPT 60
QY 51 LVTALSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSGVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSGVOLA 180
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSGVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSGVOLA 180
QY 131 DHYQONTPIGDGVPVLLPDNHNHLSQALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQONTPIGDGVPVLLPDNHNHLSQALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSGVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSGVOLA 180
QY 181 DHYQONTPIGDGVPVLLPDNHNHLSQALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHNHLSQALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Ilopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSGVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSGVOLA 180
QY 181 DHYQONTPIGDGVPVLLPDNHNHLSQALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHNHLSQALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```

```

; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match      98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Li, Xiangiang
; APPLICANT: Kain, Steve
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match      98.8%; Score 1258; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 3.4e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CIP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match      98.8%; Score 1258; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 3.4e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match      98.8%; Score 1258; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 3.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
```

Db 1 MVSKEELFTGVVPILVELDGVNCHKFSVSGEGEDATYVKLTILKFICTTGKLPVPWPT 60
Qy 51 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 191 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 191 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 9

US-09-172-053-21

; Sequence 21, Application US/09172063

; Patent No. 6150176

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.

; APPLICANT: Remington, S. James

; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR

; FILE REFERENCE: 07257/071001

; CURRENT APPLICATION NUMBER: US/09/172,063

; CURRENT FILING DATE: 1998-10-13

; EARLIER APPLICATION NUMBER: 09/094,359

; EARLIER FILING DATE: 1998-06-09

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 21

; LENGTH: 323

; TYPE: PRT

; ORGANISM: Aequorea victoria

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (0)...(0)

; OTHER INFORMATION: GT-EGFP

; US-09-172-053-21

Query Match 98.8%; Score 1258; DB 3; Length 323;

Best Local Similarity 98.7%; Pred. No. 4.2e-127;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGVNCHKFSVSGEGEDATYVKLTILKFICTTGKLPVPWPT 60
Db 35 MVSKEELFTGVVPILVELDGVNCHKFSVSGEGEDATYVKLTILKFICTTGKLPVPWPT 144
Qy 51 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 205 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 264
Qy 191 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 255 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21

; Sequence 21, Application US/09602641

; Patent No. 6608189

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
; US-09-602-641-21

Query Match 98.8%; Score 1258; DB 4; Length 323;

Best Local Similarity 98.7%; Pred. No. 4.2e-127;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGVNCHKFSVSGEGEDATYVKLTILKFICTTGKLPVPWPT 60
Db 85 MVSKEELFTGVVPILVELDGVNCHKFSVSGEGEDATYVKLTILKFICTTGKLPVPWPT 144
Qy 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 205 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 264
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6

; Sequence 6, Application US/09085305

; Patent No. 6191269

; GENERAL INFORMATION:

; APPLICANT: Pollock, Allan

; APPLICANT: Lovett, David H.

; APPLICANT: Turck, Johanna

; TITLE OF INVENTION: Selective Induction of Apoptosis in

; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal

; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bozicevic & Reed, LLP

; STREET: 285 Hamilton Ave, Suite 200

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94301

; COMPUTER READABLE FORM:

; COMPUTER TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/085,305

; FILING DATE: 29-MAY-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

Qy	181	DHYQONTPTGDG	PVLLPDNHHYSTQSALS	KDPNEKRDHMLLGFVTAAGITLGMDEL	YK	239
Db	321	DHYQONTPTGDG	PVLLPDNHHYSTQSALS	KDPNEKRDHMLLGFVTAAGITLGMDEL	YK	379
RESULT 13						
US-09-800-170-48						
; Sequence 48, Application US/09800170						
; Patent No. 6481667						
; GENERAL INFORMATION:						
; APPLICANT: Kinsella, Todd						
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES						
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK						
; CURRENT APPLICATION NUMBER: US/09/800,170						
; CURRENT FILING DATE: 2001-08-28						
; PRIOR APPLICATION NUMBER: US 60/187,130						
; PRIOR FILING DATE: 2000-03-06						
; NUMBER OF SEQ ID NOS: 90						
; SOFTWARE: Patent in version 3.1						
; SEQ ID NO 48						
; LENGTH: 434						
; TYPE: PRT						
; ORGANISM: Synechocystis PCC6803						
US-09-800-170-48						
Query Match 98.8%; Score 1258; DB 4; Length 434;						
Best Local Similarity 98.7%; Pred. No. 6.6e-127;						
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps						
Qy	1	MVSKGBELFTGV	VPILVELDGVNGHKFSVSGEGSDATY	GKLTUKFTCTTGKLPVPWPT	60	
Db	196	MVSKGBELFTGV	VPILVELDGVNGHKFSVSGEGSDATY	GKLTUKFTCTTGKLPVPWPT	25	
Qy	61	LVLTALSGVQC	FSRYPDHMKQHDFFKSAMPEGYVQ	ERTIFFKDDGNYKTRAEVKPEGDTL	12	
Db	256	LVTTLYGQC	FSRYPDHMKQHDFFKSAMPEGYVQ	ERTIFFKDDGNYKTRAEVKPEGDTL	31	
Qy	121	VNRIELKGIDF	KEDGNILGHKLEYNNSHN	VYIMADKQNGIKVNFKIRHNIEDGSVOLA	18	
Db	316	VNRIELKGIDF	KEDGNILGHKLEYNNSHN	VYIMADKQNGIKVNFKIRHNIEDGSVOLA	374	
Qy	181	DHYQONTPTGDG	PVLLPDNHHYSTQSALS	KDPNEKRDHMLLGFVTAAGITLGMDEL	YK	239
Db	376	DHYQONTPTGDG	PVLLPDNHHYSTQSALS	KDPNEKRDHMLLGFVTAAGITLGMDEL	YK	434
RESULT 14						
US-09-417-197-127						
; Sequence 127, Application US/09417197						
; Patent No. 6518021						
; GENERAL INFORMATION:						
; APPLICANT: Ole Thastrup, et al.						
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relat						
; TITLE OF INVENTION: On A Cellular Response						
; FILE REFERENCE: 3759-0110P						
; CURRENT APPLICATION NUMBER: US/09/417,197						
; CURRENT FILING DATE: 1999-10-07						
; NUMBER OF SEQ ID NOS: 143						
; SOFTWARE: Patent in version 3.0						
; SEQ ID NO 127						
; LENGTH: 442						
; TYPE: PRT						
; ORGANISM: Artificial Sequence						
; FEATURE:						
; OTHER INFORMATION: EGFP-RhoA fusion						
US-09-417-197-127						
Query Match 98.8%; Score 1258; DB 4; Length 442;						
Best Local Similarity 98.7%; Pred. No. 6.8e-127;						
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps						
Qy	1	MVSKGBELFTGV	VPILVELDGVNGHKFSVSGEGSDATY	GKLTUKFTCTTGKLPVPWPT	60	

Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-11
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170

Query Match 98.8%; Score 1258; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 7.2e-127; Indels 0; Gaps 0;
Matches 236; Conservative 1; Mismatches 2;
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:01
Job time : 13.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds
(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-64A

Perfect score: 1273

Sequence: 1 MYSKGEELFTGVVPILVELD.....VLLGFTVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1269	99.7	239	9	US-09-887-784-4
2	1269	99.7	239	12	US-10-296-953-4
3	1269	99.7	363	14	US-10-270-223-6
4	1269	99.7	893	14	US-10-257-909A-30
5	1269	99.7	1132	14	US-10-257-909A-32
6	1261	99.1	239	9	US-09-887-784-2
7	1261	99.1	239	12	US-10-296-953-2
8	1258	98.8	239	9	US-09-920-922-2
9	1258	98.8	239	9	US-09-959-745-4
10	1258	98.8	239	10	US-09-866-538-4
11	1258	98.8	239	10	US-09-797-496B-2
12	1258	98.8	239	10	US-09-794-308-4
13	1258	98.8	239	10	US-09-865-291-4
14	1258	98.8	239	12	US-10-457-982-3
15	1258	98.8	239	14	US-10-121-258-13

16	1258	98.8	239	14	US-10-221-461-7	Sequence 7, Appli
17	1258	98.8	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1258	98.8	239	14	US-10-177-390-2	Sequence 2, Appli
19	1258	98.8	239	14	US-10-338-411-3	Sequence 3, Appli
20	1258	98.8	239	15	US-10-370-570-4	Sequence 4, Appli
21	1258	98.8	239	15	US-10-389-640-3	Sequence 3, Appli
22	1258	98.8	259	14	US-10-314-861-11	Sequence 11, Appli
23	1258	98.8	281	12	US-09-931-232-1	Sequence 1, Appli
24	1258	98.8	288	14	US-10-314-861-37	Sequence 37, Appli
25	1258	98.8	293	14	US-10-314-861-35	Sequence 35, Appli
26	1258	98.8	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1258	98.8	295	14	US-10-314-861-39	Sequence 39, Appli
28	1258	98.8	299	14	US-10-314-861-33	Sequence 33, Appli
29	1258	98.8	305	14	US-10-314-861-31	Sequence 31, Appli
30	1258	98.8	308	14	US-10-033-717-35	Sequence 35, Appli
31	1258	98.8	311	14	US-10-314-861-29	Sequence 29, Appli
32	1258	98.8	320	14	US-10-338-411-11	Sequence 11, Appli
33	1258	98.8	320	15	US-10-389-640-11	Sequence 11, Appli
34	1258	98.8	323	12	US-10-457-982-21	Sequence 21, Appli
35	1258	98.8	323	14	US-10-338-411-7	Sequence 7, Appli
36	1258	98.8	323	14	US-10-338-411-13	Sequence 13, Appli
37	1258	98.8	323	15	US-10-389-640-7	Sequence 7, Appli
38	1258	98.8	323	15	US-10-389-640-13	Sequence 13, Appli
39	1258	98.8	324	14	US-10-314-861-16	Sequence 16, Appli
40	1258	98.8	345	14	US-10-338-411-5	Sequence 5, Appli
41	1258	98.8	345	15	US-10-389-640-5	Sequence 5, Appli
42	1258	98.8	346	14	US-10-338-411-9	Sequence 9, Appli
43	1258	98.8	346	15	US-10-389-640-9	Sequence 9, Appli
44	1258	98.8	359	14	US-10-033-717-33	Sequence 33, Appli
45	1258	98.8	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US2002017189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequoria Victoria
US-09-887-784-4

Query Match 99.7%; Score 1269; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 5.8e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MYSKGEELFTGVVPILVELDGVNGHKFSVSGEGDATYGKLT	KFKICTTGKLPVPWPT 60
Db	1	MYSKGEELFTGVVPILVELDGVNGHKFSVSGEGDATYGKLT	KFKICTTGKLPVPWPT 60
Qy	61	LVTALSYGVQCFSRYPDHMKQDFFKASMPGVVQERTIFFKDD	GNKYKTRAEVKFEGDTL 120
Db	61	LVTTLTSYGVQCFSRYPDHMKQDFFKASMPGVVQERTIFFKDD	GNKYKTRAEVKFEGDTL 120
Qy	121	VNRIELKGIIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKV	NFKIRHNIEDSGVQLA 180
Db	121	VNRIELKGIIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKV	NFKIRHNIEDSGVQLA 180
Qy	181	DHYQNTPTIGDGVLLPDNHYLSTQSALSKDNEKRDHMLLGFT	VTAAAGITLGMDELYK 239
Db	181	DHYQNTPTIGDGVLLPDNHYLSTQSALSKDNEKRDHMLLGFT	VTAAAGITLGMDELYK 239

RESULT 2

```

US-10-296-953-4
; Sequence 4, Application US/10286953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

```

Query Match	99.7%;	Score 1269;	DB 12;	Length 239;
Best Local Similarity	99.6%;	Pred. No. 5.8e-124;		
Matches 239;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	MVSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTKLKFICTTGKLPVPWPT 60		
DB	1	MVSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTKLKFICTTGKLPVPWPT 60		
QY	61	LVLTALSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYNTRAEVFEKDTL 120		
DB	61	LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYNTRAEVFEKDTL 120		
QY	121	VNRIELKGIDFKEDGNLGHKLEIYNYNSHNVYINADKKNGIKYKFNKIRHNIEDGSVQLA 180		
DB	121	VNRIELKGIDFKEDGNLGHKLEIYNYNSHNVYINADKKNGIKYKFNKIRHNIEDGSVQLA 180		
QY	191	DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHWMLLGFVTAAGITLGMDELYK 239		
DB	191	DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHWMLLGFVTAAGITLGMDELYK 239		

RESULT 3

```

US-10-270-223-6
; Sequence 5, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: BioImage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD
; TITLE OF INVENTION: INTACT LIVING CELLS
; TITLE OF INVENTION: INTERACTIONS BY FLOW CYTOMETRY
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: prt
; ORGANISM: Aequoria Victoria and Human
US-10-270-223-6

```

Query Match	99.7%	Score 1269	DB 14	Length 363
Best Local Similarity	99.6%	Pred. No. 1.1e-123		
Matches 238: Conservative	0	Mismatches 1	Indels 0	Gaps 0

Qy	1	MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYCKLTLKFICTTGTGKLPVPWPT	60
Db	1	MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYCKLTLKFICTTGTGKLPVPWPT	60
Qy	61	LVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIIFPKDDGNVKTAEVKFEGDIL	120
Db	61	LVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIIFPKDDGNVKTAEVKFEGDIL	120
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNYIIMADKQKNGIKVNFKIRHNIEDGSVQLA	180
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNYIIMADKQKNGIKVNFKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPTGDGPVLLPDNNHLYSTQSALSKDPNEKRDMVLLGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPTGDGPVLLPDNNHLYSTQSALSKDPNEKRDMVLLGFVTAAGITLGMDELYK	239

RESULT 4

```

US-10-257-909A-30
Sequence 30, Application US/10257909A
Publication No. US20030187056A1
GENERAL INFORMATION:
APPLICANT: Bernard R. TERRY et al.
TITLE OF INVENTION: Live cell procedures for
TITLE OF INVENTION: distribution of phospho
FILE REFERENCE: 3759-0125P
CURRENT APPLICATION NUMBER: US/10/257,909A
CURRENT FILING DATE: 2002-10-17
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 30
LENGTH: 893
TYPE: PRI
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Fusion between Aequore
US-10-257-909A-30

```

Query Match	Score 1269;	DB 14;	Length 893;
Best Local Similarity	99.6%;		
Pred. No.	3.8e-123;		
Matches 238;	Conservative	0;	Mismatches 1;
Indels 0;	Gaps	0;	

[illegible]

RESULT 5

```

US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures t
; TITLE OF INVENTION: distribution of phosph
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Fast-SEQ for Windows Version 3.0

```



```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.7%; Score 1269; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 5.4e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
DB 894 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 953
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180.
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHMLGFTVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHMLGFTVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          99.1%; Score 1261; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 4e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHMLGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHMLGFTVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          98.8%; Score 1258; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 8.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
```


APPLICANT: TSJEN, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match 98.8%; Score 1258; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 8.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGTGKLPVPWPT 60
QY 61 LVTALSYGQCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNYSQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNYSQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
Sequence 4, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSJEN, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIO-METRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REGEN1550
CURRENT APPLICATION NUMBER: US/09/865,291
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match 98.8%; Score 1258; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 8.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGTGKLPVPWPT 60
QY 61 LVTALSYGQCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHNYSQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNYSQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US20030212265A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Atsushi
APPLICANT: Llopis, Juan
APPLICANT: Wachter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR APPLICATION NUMBER: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 98.8%; Score 1258; DB 12; Length 239;
Best Local Similarity 98.7%; Pred. No. 8.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGTGKLPVPWPT 60
QY 61 LVTALSYGQCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNYSQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNYSQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 15
US-10-121-258-13
Sequence 13, Application US/10121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP1
CURRENT APPLICATION NUMBER: US/10/121,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-238-13

Query Match      98.8%; Score 1258; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. NO. 8.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
   |||||
Db 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
   |||||

QY 51 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||
Db 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNPKIRHNIEDGSVQLA 180
   |||||
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNPKIRHNIEDGSVQLA 180
   |||||

QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
   |||||
Db 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
   |||||
```

Search completed: June 21, 2004, 16:09:26
Job time : 35.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds
(without alignments)
2224.817 Million cell updates/sec

Title: US-09-887-784-64A
Perfect score: 1273
Sequence: 1 MVSKGEELFTGVVPILVELD.....VLLGFVTAAGITLGMDELYK 239
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues 283366
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1236	97.1	238	1 JQ1514	green-fluorescent
2	105	8.2	785	2 H72228	hypothetical prote
3	92.5	7.3	655	2 D83917	DNA topoisomerase
4	91.5	7.2	370	2 E70390	iron-sulfur cofact
5	91.5	7.2	861	2 H64102	leucine-tRNA ligas
6	89.5	7.0	887	2 E82590	leucyl-tRNA synthe
7	89	7.0	439	2 JH0414	synaptogamin o-p65
8	87.5	6.9	860	2 AC0582	leucyl-tRNA synthe
9	87.5	6.9	2573	2 D71614	hypothetical prote
10	87	6.8	578	1 I40794	dihydrolipoamide d
11	86.5	6.8	797	2 JC4078	protective surface
12	86.5	6.8	808	2 F64102	cellulase (EC 3.2.
13	86.5	6.8	941	2 S29043	cellulase (EC 3.2.
14	86	6.8	357	2 G81355	tRNA (uracil-5)-m
15	85.5	6.7	788	1 JDVLHH	DNA-directed DNA p
16	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
17	85	6.7	281	2 AD2052	hypothetical prote
18	85	6.7	632	2 T06586	DNA-binding protei
19	84.5	6.6	425	2 C97354	hypothetical prote
20	84.5	6.6	613	2 A99552	oligodeopeptidase
21	84.5	6.6	740	2 G95153	neuraminidase, pro
22	83.5	6.6	836	1 JDVLID	DNA-directed DNA p
23	83.5	6.6	1134	2 A60234	IgA Fc receptor pr
24	83.5	6.6	1164	1 FCSOAG	IgA Fc receptor pr
25	83	6.5	461	2 T06936	photosystem II chl
26	83	6.5	471	2 T27856	hypothetical prote
27	83	6.5	774	2 T39539	alpha-amylase homo
28	82.5	6.5	2222	1 A36028	DNA-directed DNA p
29	82	6.4	353	2 E84941	imidazoleglycerol-

ALIGNMENTS

RESULT 1

JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:1347277
A:Accession: JQ1514
A:Molecule type: DNA
A:Residues: 1-107, 'S', 109-238 <PRA1>
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99, 'P', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 'O', 26-156, 'P', 158-171, 'K', 173-238 <INO>
A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384
R:Watkins, J.N.; Campbell, A.K.
submitted to the EMBL data Library, January 1995
A:Reference number: S51330
A:Molecule type: mRNA
A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009
A:Experimental source: clone gfp1
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24, 'O', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 209-238 <PRA5>
A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011
A:Experimental source: clone gfp2
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65692; PDB:IGFL
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-91; 'Q', 92-100, 'L', 101-110, 'S', 111-120, 'T', 121-130, 'Y', 131-140, 'H', 141-150, 'D', 151-160, 'E', 161-170, 'K', 171-180, 'R', 181-190, 'G', 191-200, 'N', 201-210, 'C', 211-220, 'P', 221-230, 'F', 231-240, 'I', 241-250, 'M', 251-260, 'Q', 261-270, 'W', 271-280, 'K', 281-290, 'R', 291-300, 'L', 301-310, 'S', 311-320, 'T', 321-330, 'Y', 331-340, 'H', 341-350, 'D', 351-360, 'E', 361-370, 'K', 371-380, 'R', 381-390, 'G', 391-400, 'N', 401-410, 'C', 411-420, 'P', 421-430, 'F', 431-440, 'I', 441-450, 'M', 451-460, 'Q', 461-470, 'W', 471-480, 'K', 481-490, 'R', 491-500, 'L', 501-510, 'S', 511-520, 'T', 521-530, 'Y', 531-540, 'H', 541-550, 'D', 551-560, 'E', 561-570, 'K', 571-580, 'R', 581-590, 'G', 591-600, 'N', 601-610, 'C', 611-620, 'P', 621-630, 'F', 631-640, 'I', 641-650, 'M', 651-660, 'Q', 661-670, 'W', 671-680, 'K', 681-690, 'R', 691-700, 'L', 701-710, 'S', 711-720, 'T', 721-730, 'Y', 731-740, 'H', 741-750, 'D', 751-760, 'E', 761-770, 'K', 771-780, 'R', 781-790, 'G', 791-800, 'N', 801-810, 'C', 811-820, 'P', 821-830, 'F', 831-840, 'I', 841-850, 'M', 851-860, 'Q', 861-870, 'W', 871-880, 'K', 881-890, 'R', 891-900, 'L', 901-910, 'S', 911-920, 'T', 921-930, 'Y', 931-940, 'H', 941-950, 'D', 951-960, 'E', 961-970, 'K', 971-980, 'R', 981-990, 'G', 991-1000, 'N', 1001-1010, 'C', 1011-1020, 'P', 1021-1030, 'F', 1031-1040, 'I', 1041-1050, 'M', 1051-1060, 'Q', 1061-1070, 'W', 1071-1080, 'K', 1081-1090, 'R', 1091-1100, 'L', 1101-1110, 'S', 1111-1120, 'T', 1121-1130, 'Y', 1131-1140, 'H', 1141-1150, 'D', 1151-1160, 'E', 1161-1170, 'K', 1171-1180, 'R', 1181-1190, 'G', 1191-1200, 'N', 1201-1210, 'C', 1211-1220, 'P', 1221-1230, 'F', 1231-1240, 'I', 1241-1250, 'M', 1251-1260, 'Q', 1261-1270, 'W', 1271-1280, 'K', 1281-1290, 'R', 1291-1300, 'L', 1301-1310, 'S', 1311-1320, 'T', 1321-1330, 'Y', 1331-1340, 'H', 1341-1350, 'D', 1351-1360, 'E', 1361-1370, 'K', 1371-1380, 'R', 1381-1390, 'G', 1391-1400, 'N', 1401-1410, 'C', 1411-1420, 'P', 1421-1430, 'F', 1431-1440, 'I', 1441-1450, 'M', 1451-1460, 'Q', 1461-1470, 'W', 1471-1480, 'K', 1481-1490, 'R', 1491-1500, 'L', 1501-1510, 'S', 1511-1520, 'T', 1521-1530, 'Y', 1531-1540, 'H', 1541-1550, 'D', 1551-1560, 'E', 1561-1570, 'K', 1571-1580, 'R', 1581-1590, 'G', 1591-1600, 'N', 1601-1610, 'C', 1611-1620, 'P', 1621-1630, 'F', 1631-1640, 'I', 1641-1650, 'M', 1651-1660, 'Q', 1661-1670, 'W', 1671-1680, 'K', 1681-1690, 'R', 1691-1700, 'L', 1701-1710, 'S', 1711-1720, 'T', 1721-1730, 'Y', 1731-1740, 'H', 1741-1750, 'D', 1751-1760, 'E', 1761-1770, 'K', 1771-1780, 'R', 1781-1790, 'G', 1791-1800, 'N', 1801-1810, 'C', 1811-1820, 'P', 1821-1830, 'F', 1831-1840, 'I', 1841-1850, 'M', 1851-1860, 'Q', 1861-1870, 'W', 1871-1880, 'K', 1881-1890, 'R', 1891-1900, 'L', 1901-1910, 'S', 1911-1920, 'T', 1921-1930, 'Y', 1931-1940, 'H', 1941-1950, 'D', 1951-1960, 'E', 1961-1970, 'K', 1971-1980, 'R', 1981-1990, 'G', 1991-2000, 'N', 2001-2010, 'C', 2011-2020, 'P', 2021-2030, 'F', 2031-2040, 'I', 2041-2050, 'M', 2051-2060, 'Q', 2061-2070, 'W', 2071-2080, 'K', 2081-2090, 'R', 2091-2100, 'L', 2101-2110, 'S', 2111-2120, 'T', 2121-2130, 'Y', 2131-2140, 'H', 2141-2150, 'D', 2151-2160, 'E', 2161-2170, 'K', 2171-2180, 'R', 2181-2190, 'G', 2191-2200, 'N', 2201-2210, 'C', 2211-2220, 'P', 2221-2230, 'F', 2231-2240, 'I', 2241-2250, 'M', 2251-2260, 'Q', 2261-2270, 'W', 2271-2280, 'K', 2281-2290, 'R', 2291-2300, 'L', 2301-2310, 'S', 2311-2320, 'T', 2321-2330, 'Y', 2331-2340, 'H', 2341-2350, 'D', 2351-2360, 'E', 2361-2370, 'K', 2371-2380, 'R', 2381-2390, 'G', 2391-2400, 'N', 2401-2410, 'C', 2411-2420, 'P', 2421-2430, 'F', 2431-2440, 'I', 2441-2450, 'M', 2451-2460, 'Q', 2461-2470, 'W', 2471-2480, 'K', 2481-2490, 'R', 2491-2500, 'L', 2501-2510, 'S', 2511-2520, 'T', 2521-2530, 'Y', 2531-2540, 'H', 2541-2550, 'D', 2551-2560, 'E', 2561-2570, 'K', 2571-2580, 'R', 2581-2590, 'G', 2591-2600, 'N', 2601-2610, 'C', 2611-2620, 'P', 2621-2630, 'F', 2631-2640, 'I', 2641-2650, 'M', 2651-2660, 'Q', 2661-2670, 'W', 2671-2680, 'K', 2681-2690, 'R', 2691-2700, 'L', 2701-2710, 'S', 2711-2720, 'T', 2721-2730, 'Y', 2731-2740, 'H', 2741-2750, 'D', 2751-2760, 'E', 2761-2770, 'K', 2771-2780, 'R', 2781-2790, 'G', 2791-2800, 'N', 2801-2810, 'C', 2811-2820, 'P', 2821-2830, 'F', 2831-2840, 'I', 2841-2850, 'M', 2851-2860, 'Q', 2861-2870, 'W', 2871-2880, 'K', 2881-2890, 'R', 2891-2900, 'L', 2901-2910, 'S', 2911-2920, 'T', 2921-2930, 'Y', 2931-2940, 'H', 2941-2950, 'D', 2951-2960, 'E', 2961-2970, 'K', 2971-2980, 'R', 2981-2990, 'G', 2991-3000, 'N', 3001-3010, 'C', 3011-3020, 'P', 3021-3030, 'F', 3031-3040, 'I', 3041-3050, 'M', 3051-3060, 'Q', 3061-3070, 'W', 3071-3080, 'K', 3081-3090, 'R', 3091-3100, 'L', 3101-3110, 'S', 3111-3120, 'T', 3121-3130, 'Y', 3131-3140, 'H', 3141-3150, 'D', 3151-3160, 'E', 3161-3170, 'K', 3171-3180, 'R', 3181-3190, 'G', 3191-3200, 'N', 3201-3210, 'C', 3211-3220, 'P', 3221-3230, 'F', 3231-3240, 'I', 3241-3250, 'M', 3251-3260, 'Q', 3261-3270, 'W', 3271-3280, 'K', 3281-3290, 'R', 3291-3300, 'L', 3301-3310, 'S', 3311-3320, 'T', 3321-3330, 'Y', 3331-3340, 'H', 3341-3350, 'D', 3351-3360, 'E', 3361-3370, 'K', 3371-3380, 'R', 3381-3390, 'G', 3391-3400, 'N', 3401-3410, 'C', 3411-3420, 'P', 3421-3430, 'F', 3431-3440, 'I', 3441-3450, 'M', 3451-3460, 'Q', 3461-3470, 'W', 3471-3480, 'K', 3481-3490, 'R', 3491-3500, 'L', 3501-3510, 'S', 3511-3520, 'T', 3521-3530, 'Y', 3531-3540, 'H', 3541-3550, 'D', 3551-3560, 'E', 3561-3570, 'K', 3571-3580, 'R', 3581-3590, 'G', 3591-3600, 'N', 3601-3610, 'C', 3611-3620, 'P', 3621-3630, 'F', 3631-3640, 'I', 3641-3650, 'M', 3651-3660, 'Q', 3661-3670, 'W', 3671-3680, 'K', 3681-3690, 'R', 3691-3700, 'L', 3701-3710, 'S', 3711-3720, 'T', 3721-3730, 'Y', 3731-3740, 'H', 3741-3750, 'D', 3751-3760, 'E', 3761-3770, 'K', 3771-3780, 'R', 3781-3790, 'G', 3791-3800, 'N', 3801-3810, 'C', 3811-3820, 'P', 3821-3830, 'F', 3831-3840, 'I', 3841-3850, 'M', 3851-3860, 'Q', 3861-3870, 'W', 3871-3880, 'K', 3881-3890, 'R', 3891-3900, 'L', 3901-3910, 'S', 3911-3920, 'T', 3921-3930, 'Y', 3931-3940, 'H', 3941-3950, 'D', 3951-3960, 'E', 3961-3970, 'K', 3971-3980, 'R', 3981-3990, 'G', 3991-4000, 'N', 4001-4010, 'C', 4011-4020, 'P', 4021-4030, 'F', 4031-4040, 'I', 4041-4050, 'M', 4051-4060, 'Q', 4061-4070, 'W', 4071-4080, 'K', 4081-4090, 'R', 4091-4100, 'L', 4101-4110, 'S', 4111-4120, 'T', 4121-4130, 'Y', 4131-4140, 'H', 4141-4150, 'D', 4151-4160, 'E', 4161-4170, 'K', 4171-4180, 'R', 4181-4190, 'G', 4191-4200, 'N', 4201-4210, 'C', 4211-4220, 'P', 4221-4230, 'F', 4231-4240, 'I', 4241-4250, 'M', 4251-4260, 'Q', 4261-4270, 'W', 4271-4280, 'K', 4281-4290, 'R', 4291-4300, 'L', 4301-4310, 'S', 4311-4320, 'T', 4321-4330, 'Y', 4331-4340, 'H', 4341-4350, 'D', 4351-4360, 'E', 4361-4370, 'K', 4371-4380, 'R', 4381-4390, 'G', 4391-4400, 'N', 4401-4410, 'C', 4411-4420, 'P', 4421-4430, 'F', 4431-4440, 'I', 4441-4450, 'M', 4451-4460, 'Q', 4461-4470, 'W', 4471-4480, 'K', 4481-4490, 'R', 4491-4500, 'L', 4501-4510, 'S', 4511-4520, 'T', 4521-4530, 'Y', 4531-4540, 'H', 4541-4550, 'D', 4551-4560, 'E', 4561-4570, 'K', 4571-4580, 'R', 4581-4590, 'G', 4591-4600, 'N', 4601-4610, 'C', 4611-4620, 'P', 4621-4630, 'F', 4631-4640, 'I', 4641-4650, 'M', 4651-4660, 'Q', 4661-4670, 'W', 4671-4680, 'K', 4681-4690, 'R', 4691-4700, 'L', 4701-4710, 'S', 4711-4720, 'T', 4721-4730, 'Y', 4731-4740, 'H', 4741-4750, 'D', 4751-4760, 'E', 4761-4770, 'K', 4771-4780, 'R', 4781-4790, 'G', 4791-4800, 'N', 4801-4810, 'C', 4811-4820, 'P', 4821-4830, 'F', 4831-4840, 'I', 4841-4850, 'M', 4851-4860, 'Q', 4861-4870, 'W', 4871-4880, 'K', 4881-4890, 'R', 4891-4900, 'L', 4901-4910, 'S', 4911-4920, 'T', 4921-4930, 'Y', 4931-4940, 'H', 4941-4950, 'D', 4951-4960, 'E', 4961-4970, 'K', 4971-4980, 'R', 4981-4990, 'G', 4991-5000, 'N', 5001-5010, 'C', 5011-5020, 'P', 5021-5030, 'F', 5031-5040, 'I', 5041-5050, 'M', 5051-5060, 'Q', 5061-5070, 'W', 5071-5080, 'K', 5081-5090, 'R', 5091-5100, 'L', 5101-5110, 'S', 5111-5120, 'T', 5121-5130, 'Y', 5131-5140, 'H', 5141-5150, 'D', 5151-5160, 'E', 5161-5170, 'K', 5171-5180, 'R', 5181-5190, 'G', 5191-5200, 'N', 5201-5210, 'C', 5211-5220, 'P', 5221-5230, 'F', 5231-5240, 'I', 5241-5250, 'M', 5251-5260, 'Q', 5261-5270, 'W', 5271-5280, 'K', 5281-5290, 'R', 5291-5300, 'L', 5301-5310, 'S', 5311-5320, 'T', 5321-5330, 'Y', 5331-5340, 'H', 5341-5350, 'D', 5351-5360, 'E', 5361-5370, 'K', 5371-5380, 'R', 5381-5390, 'G', 5391-5400, 'N', 5401-5410, 'C', 5411-5420, 'P', 5421-5430, 'F', 5431-5440, 'I', 5441-5450, 'M', 5451-5460, 'Q', 5461-5470, 'W', 5471-5480, 'K', 5481-5490, 'R', 5491-5500, 'L', 5501-5510, 'S', 5511-5520, 'T', 5521-5530, 'Y', 5531-5540, 'H', 5541-5550, 'D', 5551-5560, 'E', 5561-5570, 'K', 5571-5580, 'R', 5581-5590, 'G', 5591-5600, 'N', 5601-5610, 'C', 5611-5620, 'P', 5621-5630, 'F', 5631-5640, 'I', 5641-5650, 'M', 5651-5660, 'Q', 5661-5670, 'W', 5671-5680, 'K', 5681-5690, 'R', 5691-5700, 'L', 5701-5710, 'S', 5711-5720, 'T', 5721-5730, 'Y', 5731-5740, 'H', 5741-5750, 'D', 5751-5760, 'E', 5761-5770, 'K', 5771-5780, 'R', 5781-5790, 'G', 5791-5800, 'N', 5801-5810, 'C', 5811-5820, 'P', 5821-5830, 'F', 5831-5840, 'I', 5841-5850, 'M', 5851-5860, 'Q', 5861-5870, 'W', 5871-5880, 'K', 5881-5890, 'R', 5891-5900, 'L', 5901-5910, 'S', 5911-5920, 'T', 5921-5930, 'Y', 5931-5940, 'H', 5941-5950, 'D', 5951-5960, 'E', 5961-5970, 'K', 5971-5980, 'R', 5981-5990, 'G', 5991-6000, 'N', 6001-6010, 'C', 6011-6020, 'P', 6021-6030, 'F', 6031-6040, 'I', 6041-6050, 'M', 6051-6060, 'Q', 6061-6070, 'W', 6071-6080, 'K', 6081-6090, 'R', 6091-6100, 'L', 6101-6110, 'S', 6111-6120, 'T', 6121-6130, 'Y', 6131-6140, 'H', 6141-6150, 'D', 6151-6160, 'E', 6161-6170, 'K', 6171-6180, 'R', 6181-6190, 'G', 6191-6200, 'N', 6201-6210, 'C', 6211-6220, 'P', 6221-6230, 'F', 6231-6240, 'I', 6241-6250, 'M', 6251-6260, 'Q', 6261-6270, 'W', 6271-6280, 'K', 6281-

Db 1.35 LEQNYGVLGPP-----EDP-----IRGYIRKAQYSYGD 163

Query
Best L

Query Match 7.2% Score 91.5; DB 2; Length 370;
F;318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match	7.2%	Score 91.5;	DB 2;	Length 370;
Best Local Similarity	25.9%	Pred. No. 4.4;		

Matches 50; Conservative 30; Mismatches 90; Indels 23; Gaps 8;
QY 4 KGBELFTGVV-----PILVELD-----GDVNGHKF-SVSGEG-----EGDATYVKGLTLKFICT 50
Db 164 KGVPLLTDAVQAGKPIELKNISYATFSGKHKAIGSGFLYSIDENVEPLIVGGQE 223
QY 51 TGKLP-----VPMPTLVTALSYGQCFSRYPDHMKQ-HDFPKSAMPEGYVQERTIFFKDD 104
Db 224 NGKRSGTENVVGLISLAKALEIIVNFSRYOEQLKRLDLFENLLLEA-LPDAQIVGKDA 282
QY 105 GNYKTRAEV---KFEQDGLVNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNG 161
Db 283 ERSPSISVVIMPFGFAEIVNKLSEGIYCTSGSACLSEGEYEPNKMRLKMGFSQEKALRM 342
QY 162 IKVNFIRHNIED 174
Db 343 VRESFGLLNKEE 355
RESULT 5
H64102
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)
N/Alternate names: leucyl-tRNA synthetase
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
C/Accession: H64102
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Cocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: H64102
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-861 <TIGR>
A/Cross-references: GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC22581.1; PID:g1573943;
C/Genetics:
A/Gene: leuS
C/Suprafamily: leucine-tRNA ligase
C/Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis
Query Match 7.2%; Score 91.5; DB 2; Length 861;
Best Local Similarity 24.1%; Pred. No. 13;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVPMPTLVTALSYGQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFPKD----- 103
Db 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQDF-----EFAQKYSLPKQVIAPLA 364
QY 104 DGNKTRAEVKPGDGLVNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADK-QKNGI 162
Db 365 DEBIDITKQAFVBEHGLVNSDEFDGKFN--DGAENG-----IADKLEKLV 408
QY 163 ---KVNFKIRH-----NIEDGVSQVLADHYVQNTPTGDSVLLPDMNHL- 202
Db 409 GKQVNYRLDWGVSQRVWGAPIPMLTLENGDVVPA-----PMEDPILLPEDVVD 461
QY 203 STQSALSADPN 213
Db 462 GVKSPINADPN 472
RESULT 6
E82590
leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)
C/Species: Xylella fastidiosa
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C/Accession: E82590
R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A/Reference number: A82515; MUID:20365717; PMID:10910347
A/Note: for a complete list of authors see reference number A59328 below
A/Accession: E82590
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-887 <SIM>
A/Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN001
A/Experimental source: strain 9a5c
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Laig
Briões, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canargo, L.E.A.; Carraro, D.M.; Carrer, H.
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A/Reference number: A59328
A/Contents: annotation
C/Genetics:
A/Gene: XF2176
C/Suprafamily: leucine-tRNA ligase
Query Match 7.0%; Score 89.5; DB 2; Length 887;
Best Local Similarity 22.2%; Pred. No. 20;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;
QY 50 TTGKLPVPMPTLVTALSYGQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNY-- 107
Db 329 TNEQLPV-WVANFVLMAYGTGAVMAVPGHDQDQEF--ANKYGLPTROVIALKEPKNQDE 385
QY 108 -----KTRAEVKPGDGLVNRLELKGIDFKEDGNILGHKLEYNVSHNVYI 153
Db 386 STWEPDVRWDYADKTR---EPE---LINSAPDGLDYQDAFEVLAERF----- 429
QY 154 MADQKNG--IKVNFKIRHNIEDGVSQVLADHYVQNTPTI-----GDGPVLLPDN 199
Db 430 ---RQGRQRRVNYRLK---DWGVSQRVWGCPFIVYPTCGAVPVPEDQLPVLIPEN 482
QY 200 -HYLSTQSALSADPN 216
Db 483 VAFSGTSGPIKTDPEWRK 500
RESULT 7
JH0414
synaptogamin o-p65-B - electric ray (Discothyrea ommata)
N/Alternate names: synaptic vesicle protein o-p65-B
C/Species: Discothyrea ommata
C/Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 13-Aug-1999
C/Accession: JH0414; PS0223
R/Wendland, B.; Miller, K.G.; Schilling, J.; Scheller, R.H.
Neuron 6, 993-1007, 1991
A/Title: Differential expression of the p65 gene family.
A/Reference number: JH0413; MUID:91273991; PMID:2054189
A/Accession: JH0414
A/Molecule type: mRNA
A/Residues: 1-439 <WEN>
A/Cross-references: GB:M64276; NID:g213110; PIDN:AAA49228.1; PID:g213111
A/Experimental source: electric organ
A/Accession: PS0223
A/Molecule type: protein
A/Residues: 'MLV', '26-34', 'XX', '194-199', 'X', '201-206', 'X', '322-332', 'D', '334-337' <WEN1>
C/Suprafamily: synaptotagmin; protein kinase C C2 region homology
C/Keywords: glycoprotein; membrane protein; synaptic vesicle
F/75-101/Domain: hydrophobic <HYD>
F/353-266/Domain: protein kinase C C2 region homology <KC2A>
F/284-399/Domain: protein kinase C C2 region homology <KC2B>
F/6,46/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
Query Match 7.0%; Score 89; DB 2; Length 439;
Best Local Similarity 19.5%; Pred. No. 8.9;
Matches 56; Conservative 46; Mismatches 93; Indels 92; Gaps 12;

QY 56 LVELDGDVNGHKFVSVEGEGDATYKGLTLKFKCTTGKLPV-PWPTLVLTALSYGV----- 69
DB 38 MNPIDTGDNSTEAGVPGEGND-VPEKLEKFMNELQKIPLPWALIAIIVSGLLLTLC 96
QY 70 -----QCSRYDPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKPEG----- 117
DB 97 CLCICKKCKCKKKKKKKGK-----KNDINMK---DVRGSGGNQDD 138
QY 118 --DTLVNRTELKIDFKEGNI--LGHKLEYNNSH----- 149
DB 139 DAETGTGTEGDEEKEEKEKLGKIOFSLDYDFOANQLTVGIIQAELPALDMGTSDFY 198
QY 150 -NYVIMADKQK-GIKVN-----FKIRHNIEDGSVQLA-----DHYQOQTFI 189
DB 199 VKVFLLPDKKKYETVQKKTNLPTFNESFVKVPYQELGGKTLMAVYDFDRFSKHD 258
QY 190 GDGPVLLPD-----NHVLTQSALSKDNEKRDEHVLGFTVTAAG 229
DB 259 GQTVLMTKVDLQQLQLEWRDLESABKEPEKLGIDICTSLRYVPTAG 305

RESULT 8
AC0582
A:Title: leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S
A:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0582
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Barry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0582
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-860 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:G16501899; GSPDB:GN00176
C:Genetics:
A:Gene: STY0699
C:Superfamily: leucine-tRNA ligase

Query Match 6.9%; Score 87.5; DB 2; Length 860;
Best Local Similarity 23.3%; Pred. No. 29;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVWPPTLVLTALSYGVQCSRYDPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKT 109
DB 314 TGEIIPV-WAANFVLMYEGTGAVMVPGH-DQD-YEFASKYGLTTKPVLAADGSEPD 370
QY 110 RAUVKEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHVIMADKQKNGIKVNFKIR 169
DB 371 SEQALTEKGVLFNSGFGLAFEAFAFNAIADKL-----AEKGVGERKVNRLR 418
QY 170 H-----NIEDGSVOLADHYQOQTFIIGDGPVLLPDNHYL-STQSALSKOP 212
DB 419 DWGVSQRVYGAIPWVTLLEDGV-----LTPEDQLPVILPVDVMDGIISPIKADP 471

RESULT 9
D71614
hypothetical protein PF0460c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: D71614
R:Gardner, M.J.; Tetelin, H.; Cummings, D.J.; Carucci, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
```

```
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: D71614
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2573 <GAR>
A:Cross-references: GB:AE001396; GB:AE001388; NID:G3845188; PIDN:AACT71881.1; PID:G384519
A:Experimental source: Clone 3D7
C:Genetics:
A:Gene: PFB0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;
Best Local Similarity 26.2%; Pred. No. 1.2e+02;
Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VOERTIFFKD--DGNVYKTRAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYN--YN 149
DB 126 LKKTETILCKDIKSGSDNDPMDIEISLFPKDDMVDDKELK--DFEKSLSKIKNKEVNFY 183
QY 150 NYVIMADKQKNGIKVNFKIRHNIEDGSVOLADHYQOQTFIIGDGPVLLPDNHYLST 209
DB 184 NLHIEKNKKKDEKKNKIHNNDENNM-----IYYKNI---DKTHYLDNNVVHILND 236
QY 210 KDPNEKRDEH 219
DB 237 TVLKRDEYM 246

RESULT 10
I40794
dihydropolipamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum
N:Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase
hydrogenase complex chain E3; S-complex 50K chain
C:Species: Clostridium magnum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: I40794
R:Kruger, N.; Opermann, F.B.; Lorenzl, H.; Steinbuechel, A.
J. Bacteriol 176, 3614-3630, 1994
A:Title: Biochemical and molecular characterization of the Clostridium magnum acetoin de
A:Reference number: I40789; MUID:94266715; PMID:8206840
A:Accession: I40794
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-578 <KRU>
A:Cross-references: GB:I31844; NID:G472324; PIDN:AAA21748.1; PID:G472330
C:Function:
A:Description: catalyzes the oxidation of dihydropolipamide to lipolipamide using NAD
A:Pathway: acetoin dehydrogenase enzyme system
C:Superfamily: Alkaligenes dihydropolipamide dehydrogenase; dihydropolipamide dehydrogenase
C:Keywords: FAD; flavoprotein; lipolipamide; NAD; oxidoreductase; redox-active disulfide
F;S-77/Domain: lipoyl/biotin-binding homology <LPS>
F;117-145/Region: beta-alpha-beta FAD nucleotide-binding fold
F;119-561/Domain: dihydropolipamide dehydrogenase homology <DLD>
F;287-315/Region: beta-alpha-beta NAD nucleotide-binding fold
F;153-158/Disulfide bonds: redox-active #status predicted

Query Match 6.8%; Score 87; DB 1; Length 578;
Best Local Similarity 23.3%; Pred. No. 19;
Matches 54; Conservative 40; Mismatches 84; Indels 52; Gaps 12;

QY 10 TGVVPIVLVDGVDNGHKFVSVEGEGDATYKGLTLK-----FICTTGKLPVWPPTLV 63
DB 255 TGSMPPIEIE---GNKLS---GVIDST-GALSLENPESITAIIGGVIGVEFASIFN 305
QY 64 ALSYGVQCSRYDPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTLVNR 123
DB 306 SLGCKVSIIEMLPHILPMDREISEI-----AKAKLIRDGININN 346
QY 124 IELKGDIDFKEDG---NILGHKLEYNNSHVIMADKQK--GIKNFKIRHNIEDGSVQ 178
DB 347 CKVTRIEQGEDGLKVFIDGKGESIDVEKVLIAVGRRSNIEGLDVE-KIGVTEGSGII 405
```



```

Db      440 IGYGTSGISVQASVKQNFGLGTGAASVSIAGTKNDYGTSVNLGYTEPYPTKDGVS LGGNV 499
Qy      100 FFKDGNKYKRAEVKFGEDTLVNRRIELKIDGDKDGNI---LGH-----KLEYNVNS 148
Db      500 FFENYDNSKSDTSSTNYKRTTYGVSNTL-GFPWNENNSYYVGLGHTYKNSIFALEYN---555
Qy      149 HNVVIMADKQK-NGIKVNFKIRHNIEDGSVQLADHYQQ-----NTPIGDGPVLL 196
Db      556 RNLVIQSMKFKNGIKTN-----DPDFSGFNWYNSLNRGYFPTKGVKASLG-GRVTI 606
Qy      197 P--DNHYLSTQSALS KDPNEKRDRHMLVLLGFVTAAGITLG 233
Db      607 PGSNKKYTKLSADVQGFPLDRDRHLNWWVSASAKASYANG 645

RESULT 13
S29043
cellulae (EC 3.2.1.4) - Bacillus sp.
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Bacillus sp.
C/Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
R/Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
J. Gen. Microbiol. 136, 1327-1334, 1990
A/Title: Molecular cloning and nucleoside sequence of a gene for alkaline cell
A/Reference number: S29043; MUID:91037937; PMID:2230718
A/Accession: S29043
A/Molecule type: DNA
A/Residues: 1-941 <OZA>
A/Cross-references: EMBL:M27420; NID:gl42664; PID:AAA22304.1; PID:gl42665
R/Shirai, T.; Yanane, T.; Hidaka, T.; Kuyama, K.; Suzuki, A.; Ashida, T.; Ozak
J. Biochem. 122, 683-685, 1997
A/Title: Crystallization and preliminary X-ray analysis of a truncated family
A/Reference number: PC4404; MUID:98060488; PMID:9399567
A/Accession: PC4404
A/Molecule type: protein
A/Residues: 228-584 <SHI>
A/Experimental source: strain KSM-635
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans
A/Pathway: cellulose degradation
C/Superfamily: Bacillus sp. KSM-635 alkaline cellulase; S-layer repeat homolog
C/Keywords: glycosidase; hydrolase; polysaccharide degradation
F/41-95/Domain: S-layer repeat homolog <SLR1>
F/101-153/Domain: S-layer repeat homolog <SLR2>
F/164-219/Domain: S-layer repeat homolog <SLR3>
F/766-908/Domain: Thermotoga xylanase A amino-terminal repeat homolog <TXA>

Query Match 6.8%; Score 86.5; DB 2; Length 941;
Best Local Similarity 19.8%; Pred. No.39;
Matches 48; Conservative 34; Mismatches 61; Indels 99; Gaps 11

Qy      16 LVELDGVNGHKFVSGEGEDATYGLTLKFICTGKLPVPWPTLVLTALS-YGVQCFSR 74
Db      240 LVELNG-----QLTLAGE---DGT-----PVQLRGMSTHGLQWFG- 271

Qy      75 YPDHMKQHDFFKSPGPGYQVQERTIPFKDDGNYKTRAEVKFEGDTLVNRIELKGDIFKED 134
Db      272 --EIVNENAFVSLNDGWSNMIRLAMYIGENGYATNPEVK---DLVYEGIELA-----319

Qy      135 GNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSV---GLADHYQONTPIG 190
Db      320 -----FEHDMYIVDWH---VHAPCDPRADVYSAYDFEESIAHYKDH-----360

Qy      191 DGPVLLPDNHYLSTQSALS KDPN-----EKRDHMYL 221
Db      361 -----PKNHYIIVELANEPSPNNNGGPLTNDKGVKAEVKEPIVEMLRKGDGMIL 414

Qy      222 LG 223
Db      415 VG 416

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-64A

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1242	97.6	238	1	GFP_AEQVI
2	91.5	7.2	861	1	SYL_HAEIN
3	89.5	7.0	879	1	SYL_XYLFA
4	89	7.0	439	1	SY62 DISOM
5	88.5	7.0	1603	1	VIT4 CAEL
6	87.5	6.9	860	1	SYL_SALTI
7	87.5	6.9	860	1	SYL_SALTY
8	86.5	6.8	533	1	CP51 CANGA
9	86.5	6.8	795	1	D152 HAEIN
10	86.5	6.8	797	1	D151 HAEIN
11	86.5	6.8	879	1	SYL_XYLFT
12	86.5	6.8	941	1	GUN_BACS6
13	86	6.8	357	1	TRMA CAMJE
14	85.5	6.7	788	1	DPOL HPBHE
15	85.5	6.7	793	1	D153 HAEIN
16	85.5	6.7	886	1	ITH3 MESAU
17	84.5	6.6	501	1	AMPA WIGBR
18	84.5	6.6	613	1	PEPF_MCPU
19	84.5	6.6	859	1	SYL_SHSON
20	84	6.6	504	1	YC03 KLEPN
21	83.5	6.6	538	1	GRBE RAT
22	83.5	6.6	658	1	ADAS HUMAN
23	83.5	6.6	1164	1	BAG STRAG
24	83	6.5	461	1	PSBC_CVAPA
25	83	6.5	774	1	AMY2 SCHPO
26	82.5	6.5	2222	1	DP0E YEAST
27	82	6.4	353	1	HIS7 BUCAI
28	82	6.4	682	1	PRC_ECOLI
29	82	6.4	689	1	AC2L HUMAN
30	82	6.4	752	1	NECI_RAT
31	82	6.4	874	1	SLAP_BACLI
32	81.5	6.4	269	1	PENK MOUSE
33	81.5	6.4	589	1	SYD_HAEDU
					P42212 aequorea vi
					P43827 haemophilus
					Q9pb98 xylella fas
					P24506 discopyge o
					P18947 caenorhabdi
					Q828h5 salmonella
					Q828z6 salmonella
					P50859 candida gla
					P44935 haemophilus
					P46024 haemophilus
					Q87c65 xylella fas
					P19424 bacillus sp
					Q9pp92 campylobact
					P13846 heron hepat
					O32629 haemophilus
					P97280 mesocricetu
					Q8d295 wigglewort
					Q98q0 mycoplasma
					Q8ehp4 shewanella
					O48449 klebsiella
					O88900 rattus norv
					O00116 homo sapien
					P27951 streptococc
					P48104 cyanophora
					O42918 schizosacch
					P21951 saccharomyc
					P57203 buchnera ap
					P23865 escherichia
					Q9nub1 homo sapien
					P28840 rattus norv
					P49052 bacillus li
					P22005 mus musculu
					Q7vnf0 haemophilus

RESULT 1
GFP_AEQVI STANDARD; PRT; 238 AA.
ID P42212; Q17104; Q27903;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92175527; PubMed=1347277;
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA Cormier M.J.;
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein.";
RL Gene 111:229-233(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94185810; PubMed=8137953;
RA Inouye S., Tsuji F.I.;
RT "Aequorea green fluorescent protein. Expression of the gene and
RT fluorescence characteristics of the recombinant protein.";
RL FEBS Lett. 341:277-280(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
RT protein by modification of its codon usage.";
RL Plant Mol. Biol. 33:989-999(1997).
RN [4]
RP CHROMOPHORE.
RX MEDLINE=93192221; PubMed=8448132;
RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;
RT "Chemical structure of the hexapeptide chromophore of the Aequorea
RT green-fluorescent protein.";
RL Biochemistry 32:1212-1218(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=96355665; PubMed=8703075;
RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,
RA Remington S.J.;
RT "Crystal structure of the Aequorea victoria green fluorescent
RT protein.";
RL Science 273:1392-1395(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=98294543; PubMed=9631087;
RA Yang F., Moss L.G., Phillips G.N. Jr.;
RT "The molecular structure of green fluorescent protein.";
RL Nat. Biotechnol. 14:1246-1251(1996).

ALIGNMENTS

34 81 6.4 336 1 YD48 METJA Q58743 methanococc
35 81 6.4 1224 1 COPA HUMAN P53621 homo sapien
36 80.5 6.3 393 1 TRMB HELPY O25443 helicobacte
37 80.5 6.3 658 1 ADAS_CAVPO P97275 cavia porce
38 80.5 6.3 860 1 SYL_ECO57 Q8xbn8 escherichia
39 80.5 6.3 860 1 SYL_ECOL6 Q8fjy9 escherichia
40 80.5 6.3 860 1 SYL_ECOLI P07813 escherichia
41 80.5 6.3 1157 1 POL_SFV3L P27401 simian foam
42 80 6.3 737 1 OPTI DROME P91679 drosophila
43 79.5 6.2 217 1 EXP1 ERWCA P33882 erwinia car
44 79.5 6.2 312 1 TRXB_CHLMU Q9pht7 chlamydia m
45 79.5 6.2 468 1 CLNA_AZOCA P94126 azorhizobiu


```
Db 181 HQOONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDMVLLFEVTAAGITHGMDELYK 238
RESULT 2
SVL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32774; AAC22581.1; -;
DR F01; H64102; H64102.
DR TIGR; H10921; -.
DR HAMAP; MF_00049; -; 1.
DR InterPro; IPR002302; Leu-TRNAsyntla.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_fleRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNT1LEU.
DR TIGRFAMs; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY)
SQ SEQUENCE 861 AA; 97750 MW; EB93304F684C8FB7 CRC64;
Query Match
Best Local Similarity 24.1%; DB 1; Length 861;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVPHPTLVLTALSYGQCFSRYPDHMKQHDFFKASMPGYGVERIIFPKD----- 103
DB 314 TGDKLPI-RWANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKSLPIKQVIAPLA 364
```

```
QY 104 DGNKYTRAEVKFEGDFTLVNRIELKGDIFKEDGNILGHKLEYNNYNNHNYIMADK-QKNGI 162
Db DEEIDLTKQAFVEHGKLVNSDFDGNF--DGAFNG-----IADKLEKLG 408
QY 163 ---KVNFKIRH-----NIEDGSVOLADHYQONTPIGDGPVLLPDNHYL- 202
Db 409 GKRQVNYRLRDMGVSQRQRYGAPIPMLTLLENGDVVPA-----PMEDLPILPEDVVD 461
QY 203 STQSALSQKDPN 213
Db 462 GVKSPINADPN 472
RESULT 3
SVL_XYLFA
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9PBG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=9a56;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach P.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Faciniani A.P., Cristofani A.J.S., Dias-Neto E., Docena C., El-Dorri H.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega P.G., Nunes L.R., Oliveira M.A., Paris A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vaglada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF004031; AAF84975.1; ALT_INIT.
DR HAMAP; MF_00049; -; 1.
```

```
DR InterPro: IPR002302; Leu-trnasynt1a.
DR InterPro: IPR002300; trna-synt-1a.
DR InterPro: IPR001412; trna-synt-1.
DR InterPro: IPR009008; Valrs-1fers_edit.
DR Pfam: PF00133; trna-synt-1; 1.
DR PRINTS: PRO0985; trnasynthLEU.
DR TIGRFAMs: TIGR00396; leus_bact; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY)
SQ SEQUENCE 879 AA; 99796 MW; 9FDCB99202919E CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 9.4;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLVPPPTLVTSYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107
DB 321 TNEQLVW-VWAFVLMAYGAGVAVPGHDDQDEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 108 -----KTRAEVKFEGDTLVNRILKGDIDFKEDGNILGHKLEYNYNHNVYI 153
DB 378 STWEPDVRWDYADKTR---EFE---LINSAEFDGLDYQDAFEVLAERPE----- 421
QY 154 MADKQKNG-IKVNFKIRHNIEDSGVOLADHYQNTPI-----GDGPVLLPDN 199
DB 422 ---RQGRQRRVNYLR-----DWGSRQRYWGCPPIVYPTCGAVPEVDQLPVILPEN 474
QY 200 -HYLSTQSALSADPNKR 216
DB 475 VAFSGTGSPKTPDEWRK 492

RESULT 4
SY62_DISOM STANDARD; PRT; 439 AA.
ID VIT4_CABEL
AC P24506;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Synaptotagmin B (Synaptic vesicle protein O-p65-B).
GN P65-B.
OS Discopyle ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hyphosqualea; Pristiogalea; Batoidea;
OC Torpediniformes; Narcinoidei; Narcinidae; Discopyle.
OX NCBI_TaxID=7785;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91273991; PubMed=2054189;
RA Wendland B., Miller K.G., Schilling J., Scheller R.H.;
RT "Differential expression of the p65 gene family.";
RL Neuron 6:993-1007(1991).
CC -1- FUNCTION: May have a regulatory role in the membrane interactions
CC during trafficking of synaptic vesicles at the active zone of the
CC synapse. It binds acidic phospholipids with a specificity that
CC requires the presence of both an acidic head group and a diacyl
CC backbone.
CC -1- SUBUNIT: Homodimer or homotrimer (possible).
CC -1- SUBCELLULAR LOCATION: Synaptic vesicles in neurons.
CC -1- TISSUE SPECIFICITY: Spinal cord, brainstem, midbrain and electric
CC organ.
CC -1- SIMILARITY: Contains 2 C2 domains.
CC -1- SIMILARITY: Belongs to the synaptotagmin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M64276; AAA49228.1; -.
DR PIR: JH0414; JH0414.
DR HSSP: P21707; 1BYN.
DR InterPro: IPR000008; C2.
DR InterPro: IPR008973; C2_CaLB.
DR InterPro: IPR002149; LRI.
DR Pfam: PF00168; C2; 2.
DR SMART: SM00239; C2; 2.
DR PROSITE: PS00499; C2_DOMAIN_1; 2.
DR PROSITE: PS00004; C2_DOMAIN_2; 2.
KW Transmembrane; Repeat; Synapse; Multigene family; Glycoprotein.
FT DOMAIN 1 74 VESICULAR (POTENTIAL).
FT TRANSMEM 75 101 POTENTIAL.
FT DOMAIN 102 439 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 153 399 PHOSPHOLIPID BINDING (PROBABLE).
FT DOMAIN 173 262 C2 DOMAIN 1.
FT DOMAIN 304 395 C2 DOMAIN 2.
FT CARBOHYD 6 6 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 46 46 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 439 AA; 49278 MW; 2033F05FD8C69F39 CRC64;

Query Match 7.0%; Score 89; DB 1; Length 439;
Best Local Similarity 19.5%; Pred. No. 4.5;
Matches 56; Conservative 46; Mismatches 93; Indels 92; Gaps 12;

QY 16 LVLEDGVNKHFSVSGEGDATYKLTLCITCTKGLPV-PWPTLVTSALSGV----- 69
DB 38 MNPIDTGDNSTAGVPEGKND-VFEKKEKFMNELQKIPLPWALIAIVSGLLLTLC 96
QY 70 -----QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEG-- 117
DB 97 CLICIKKCKCKKKKKKGGK-----KNDINMK---DVKSGSGNQDDD 138
QY 118 --DTLVNRIELKGDIDFKEDGNI--LGHKLEYNYNH-- 149
DB 139 DAETGLTEGEDKEEAEKEEKLKIQFLSDYDFQANLVGIIQAAELPALDMGTSDPY 198
QY 150 -NVYIMADKQKN-GIKVN-----FKIRHNIEDSGVOLA-----DHVQNTPI 189
DB 199 VKVFLPLDPKKKYETKVQKTLNPTNFSFVKVPYQELGGKTLMAVVDFFRSKDCI 258
QY 190 GDGPVLLPD-----NHYLSTQSALSADPNKRKDHMLVIGFVTAAG 229
DB 259 GQVTVLTMTKVDLQQLLEWRDLSEAKPEKLGDICTSLRYVPTAG 305

RESULT 5
VIT4_CABEL STANDARD; PRT; 1603 AA.
ID VIT4_CABEL
AC P18947; Q9BPP3.
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
GN Vit-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T., Spieth J., Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 1-71 FROM N.A.
RP
```

```
RX MEDLINE=95269643; PubMed=4022780;
RA Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;
RT "The C. elegans vitellogenin genes: short sequence repeats in the
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -1- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
CC building the intestine of adult hermaphroditic individuals; they
CC are cotranslationally secreted into the body cavity and
CC subsequently taken up by the gonad.
CC -1- SIMILARITY: Contains 1 WVD domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC024137; AAK09074.1; -.
DR EMBL; M11498; AAA28163.1; -.
DR EMBL; X02754; CAA26531.1; -.
DR PIR; A43084; A43084.
DR WormPep; F5908.2; CE26817.
DR InterPro; IPR001747; Lipid transprt_N.
DR InterPro; IPR001846; WVD_D.
DR Pfam; PF01347; Vitellogenin_N; 1.
DR SMART; SM00638; LPD_N; 1.
DR SMART; SM00216; WVD; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 15
FT CHAIN 16 1603
FT DOMAIN 1308 1455
FT CONFLICT 30 30
FT CONFLICT 169 169
FT CONFLICT 183 187
FT CONFLICT 275 275
SQ SEQUENCE 1603 AA; 186307 MW; E303170325BC99BB CRC64;

Query Match 7.0%; Score 88.5; DB 1; Length 1603;
Best Local Similarity 23.4%; Pred. No. 23;
Matches 52; Conservative 33; Mismatches 69; Indels 69; Gaps 12;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGVKTLKPICTTGKLPVWPPT 60
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
162 MESDKDLSLFPNVHKTMEGDCEV---AYTIVQEG-GKTIYTKSVNFDKCITR-----PE 211
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKFSAMPEG-VVQERTIF---FKDDG----- 105
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
212 TAYGLRFGSGC-----KECEKEGFQVQPTVYTFYTFKNEKLQSEVNSIYT 257
QY 106 -----NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHVYIMAD 156
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
258 LNVNGQEVVKSETRAKVTFVEESKINR-EIK-----KVSQKKEIYVSMENKLEIQ 308
QY 157 KQKNG-----IKVNFKIRHNIEDGSVQLADHYQQNTP 188
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 FYKQGDKAENVPFKAIEIEQKV-EQLEEIPFQIEH-EQNTP 348

RESULT 6
ID SYL SALTII STANDARD; PRT; 860 AA.
AC Q828H5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR STY0699 OR T2219.
OS Salmonella typhi.
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=215334947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodymani V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL627267; CAD05125.1; -.
DR EMBL; AE016841; AA069822.1; -.
DR HAMAP; MF 00049; -.
DR InterPro; IPR002302; Leu-tRNAsyntla.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_IleRS_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "RMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
FT SEQUENCE 860 AA; 96940 MW; 2F95E480BBAB23C4 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVPWPPTLVLTALSYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFFDGNGYKT 109
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 TGEIIPV-WAANFVLMYGTGVMAVPGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKVNFKIR 169
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 SEQALTEKGVLFNSGEGFDGLAFAEAFNAIDKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NIEDGSVQLADHYQQNTPITGDGVLLPDNHYL-STQSALSKDP 212
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
419 DNGVSRQRYWGAPIPMVLTLEDGTV-----LPTPEDQLPVILPDVVDGITSPKADP 471
```

```
RESULT 7
SYL_SALT
ID SYL_SALT STANDARD; PRT; 860 AA.
AC Q8ZQ26;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR STM0648;
OS Bacteroides typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diposphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AE008725; AAL19599.1; -.
CC StyGene; SG????; leus.
CC HAMAP; MF 00049; -.
CC InterPro; IPR002302; Leu-TRNAsyntla.
CC InterPro; IPR002300; tRNA-synt la.
CC InterPro; IPR001412; tRNA-synt I.
CC InterPro; IPR009008; ValRS lIeRS_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMS; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 42 52 "HIGH" REGION.
CC SITE 619 623 "KMSKS" REGION.
CC BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;

Query Match 6.98; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVWPMTLVTALSYGVQCFSRYPDHMKQDFFKSAPEGVQVORTIFFKDDGNYKT 109
Db 314 TGEIPV-WAANFVLMYEGTCAWVPGH-DQRD-YEFASKYGLTIKPVLAADGSEPD 370
QY 110 RAEVKFGDGLVNRLEKIDFKEDGNILGHKLEYNYNHNVMADKQKNGIKVNFKIR 169
Db 371 SEQALTEKGVLFNSGSEFDGLAFAFAFNAIAKLV-----AEKGVGERKVNRLR 418
QY 170 H-----NIEDGSQVLADHYQQNTPIGDGPVLAPDNHYL-STOSALSQDP 212
Db 419 DWGVSQRQYWGAPIPVMTLEDGTV-----LTPEDQLPVLPEDVMDGITSPIKADP 471

RESULT 8
CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYP11) (P450-LIA1) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RA "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility."
RL Antimicrob. Agents Chemother. 39:2708-2717 (1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RA "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIA1) gene fragment."
RL J. Clin. Microbiol. 32:1902-1907 (1994).
CC -1- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: Obtusifolitol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -1- PATHWAY: Ergosterol biosynthesis.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; L40389; AAB02329.1; -.
CC ENBL; S75389; AAB32679.1; -.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
CC Sterol biosynthesis; NADP.
CC METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC CONFLICT 64 64 I -> M (IN REF. 2).
CC CONFLICT 473 473 I -> T (IN REF. 2).
CC SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.8%; Pred. No. 9.1;
Matches 45; Conservative 32; Mismatches 80; Indels 49; Gaps 9;

QY 25 GHKFSVS---GEGEQATYKGLTLKFCITGKLPVWPMTLVTALSYGVQCFSRYPDH--M 79
```



```
Db 427 IGYTGESIQASVSKODNFGTGAASVSIAGTKNDYGTSVNLGYTEPYFTKDGVSIGNV 486
QY 100 FFKDDGNYKRAEVKFGDTLVNRILKGDHFKEDGNI---LGH-----KLEYNYS 148
Db 437 FFENYDMSKSDTSNRYKRTYTGNSVTI-GFPVNNNSYYVGLGHTYKNSNFALEYN--- 542
QY 149 HNYVIWADKOK-NGIKVNFKIRHNIEDSGVOLADHYQQ-----NTPIGDGPVLL 196
Db 543 RNLVIQSMKFGNGIKTN-----DFDFSFGWYNSLNRGYFPTKGVKASLG-GRVTI 593
QY 197 P--DNHLYSTQSALSADPNKRDHVMULLGFTVTAAGITLG 233
Db 594 PGSDNRYKYLSDVQGFYPLDRDHLWVYSKASAGYANG 632

RESULT 11
SYL_XYLFT
ID _SYL_XYLFT STANDARD; PRT; 879 AA.
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS_O3 PD1230.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Slyke M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Feilille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasakii F.T., Sena J.A.D.,
RA de Souza A.D., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa.";
RL J. Bacteriol. 185:1018-1026(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99823 MW; 4C2EE01B8FDC497E CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 17;
Matches 44; Conservative 28; Mismatches 69; Indels 57; Gaps 10;

QY 50 TTCKLPVPWPTLTALSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107
Db 321 TNEQLPV-WVANFVLMAYGTGAVMAVPGHDQDQEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 108 -----KTRAEVKFEGDTLVNRILKGDHFKEDGNIHGKLEYNNSHNVIY 153
Db 378 STWEPDVWRDWDYADKTR---EFE---LINSAFDGLDYQGAFEVLAERF----- 421
QY 154 MADKQKNG-IKVNFKIRHNIEDSGVOLADHYQQNTPI-----GDGPVLLPON 199
Db 422 ---RQGRGQRRVNYRLR---DWGVSQRQYWGCPDVIVYCTCGAVPVPENQLPVLIPEN 474
QY 200 -HYLSTQSALSADPNKNEKR 216
Db 475 VAFSGTGSPDKTDPWRK 492

RESULT 12
GUN_BACS6
ID _GUN_BACS6 STANDARD; PRT; 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (alkaline cellulase).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=2230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT "Molecular cloning and nucleotide sequence of a gene for alkaline
RT cellulase from Bacillus sp. KSM-635.";
RL J. Gen. Microbiol. 136:1327-1334(1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of Glycosyl
CC hydrolases).
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds
(without alignments)
2458.984 Million cell updates/sec

Title: US-09-887-784-64A
Perfect score: 1273
Sequence: 1 MVSGEELFTGVVPILVELD.....VLLGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTRMBL_25:*
- 1: sp_archea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_prodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1238	97.3	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1235	97.0	238	5 Q93125	Q93125 aequorea vi
3	1233	96.9	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1232	96.8	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1200	94.3	238	5 Q17105	Q17105 aequorea vi
6	1185	93.1	238	5 Q17106	Q17106 aequorea vi
7	1080	84.8	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1076	84.5	238	5 Q8WPD95	Q8wtp95 aequorea ma
9	1072	84.2	238	5 Q8WTC4	Q8wtc4 aequorea ma
10	1070	84.1	238	5 Q8WTD0	Q8wtcd0 aequorea ma
11	1069	84.0	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1069	84.0	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1067	83.8	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1065	83.7	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	256.5	20.1	225	5 Q95UA7	Q95ua7 montastraea
16	256.5	20.1	225	5 Q7Z0W5	Q7z0w5 montastraea

17	251	19.7	225	5	Q963F5	Q963f5 montastraea
18	248.5	19.5	236	5	Q8T6U0	Q8t6u0 dendronephth
19	244	19.2	225	5	Q8I6J8	Q8i6j8 trachyphyll
20	242.5	19.0	225	5	Q7Z0W9	Q7z0w9 montastraea
21	242.5	19.0	266	5	Q9U6Y3	Q9u6y3 clavularia
22	237	18.6	225	5	Q7Z0W4	Q7z0w4 montastraea
23	236	18.5	224	5	Q8MU48	Q8mu48 montastraea
24	232	18.2	225	5	Q8T5F1	Q8t5f1 montastraea
25	216.5	17.0	259	5	Q8MMA2	Q8mma2 agaricia fr
26	216	17.0	239	5	Q8MMA1	Q8mma1 agaricia ag
27	214	16.8	227	5	Q7Z0W6	Q7z0w6 montastraea
28	214	16.8	234	5	Q7Z0W7	Q7z0w7 montastraea
29	209	16.4	234	5	Q8T5F2	Q8t5f2 montastraea
30	209	16.4	234	5	Q8MU47	Q8mu47 montastraea
31	208.5	16.4	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	207.5	16.3	238	5	Q9BLV9	Q9blv9 renilla mue
33	206	16.2	227	5	Q962P9	Q962p9 montastraea
34	206	16.2	227	5	Q7Z0W8	Q7z0w8 montastraea
35	205.5	16.1	232	5	Q9GP15	Q9gp15 anemonia su
36	204	16.0	221	5	Q95P04	Q95p04 goniotopora t
37	202.5	15.9	222	5	Q7Z168	Q7z168 ceriantanthus
38	202.5	15.9	225	5	Q8T6T9	Q8t6t9 radiantanthus
39	202	15.9	227	5	Q95VT0	Q95vt0 montastraea
40	202	15.9	235	5	Q8T5F0	Q8t5f0 scolymlia cu
41	201.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
42	201.5	15.8	232	5	Q9GZ28	Q9gz28 anemonia s
43	199.5	15.7	214	5	Q86LV7	Q86lv7 meandrina m
44	198.5	15.6	214	5	Q86LV8	Q86lv8 meandrina m
45	197.5	15.5	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	ID	Q8GHE2	PRELIMINARY;	PRT;	238 AA.
AC	Q8GHE2;				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Green fluorescence protein.				
GN	2289GFP.				
OS	Azotobacter vinelandii.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Azotobacter.				
OX	NCBI_TaxID=354;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DSM2289;				
RA	Koranyi P., Berenyi M., Burg K.;				
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria				
RT	Azomonas and Azotobacter."				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF324408; AAN86140.1; -.				
DR	GO; GO:0006091; P:energy pathways; IEA.				
DR	InterPro; IPR009017; GFP like.				
DR	InterPro; IPR000786; Green_fl_protein.				
DR	Fram; PF01353; GFP; 1.				
DR	PRINTS; PR01229; GFLUORESCENT.				
DR	ProDom; PD013756; Green fl protein; 1.				
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;				

Query Match	97.3%;	Score 1238;	DB 2;	Length 238;
Best Local Similarity	97.5%;	Pred. No. 1.2e-96;		
Matches 232;	Conservative 2;	Mismatches 4;	Indels 0;	Gaps 0;
Qy	2	VSKGEELFTGVVPILVELDGVNGHKFSVSGEGGDATYGLTKFKICTTGKLPVPMPTL	61	
Db	1	MSKGEELFTGVVPILVELDGVNGHKFSVSGEGGDATYGLTKFKICTTGKLPVPMPTL	60	
Qy	62	VTALSYGVOCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV	121	

```

Db      61 VITFSYGVQCFRYPDHMKRHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY      122 NRIELKGIIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVOLAD 181
Db      121 NRIELKGIIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVOLAD 180
QY      132 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
Db      131 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 238

RESULT 2
Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,
RA Brown A.J.P.;
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESCENT.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 97.0%; Score 1235; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 2.2e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYKGLTKLKFICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYKGLTKLKFICTTGKLPVWPPTL 60

QY 62 VTALSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Db 61 VITFSYGVQCFRYPDHMKRHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVOLAD 181
Db 121 NRIELKGIIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 238

RESULT 3
Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
  Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324405; AAN86137.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESCENT.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.9%; Score 1233; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 3.3e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYKGLTKLKFICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYKGLTKLKFICTTGKLPVWPPTL 60

QY 62 VTALSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Db 61 VITFSYGVQCFRYPDHMKRHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVOLAD 181
Db 121 NRIELKGIIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 238

RESULT 4
Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
  Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESCENT.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.8%; Score 1232; DB 2; Length 238;

```

```
Best Local Similarity 97.1%; Pred. No. 4e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIILVELDGVNKHGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGELFTGVVPIILVELDGVNKHGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTALSYGVCFSRYPDHMKQHDFFKSPAMPEGYVQERTIFFKDDGNKTKRAEVKFECDTLV 121
DB 61 VTTFSYGVCFSRYPDHMKRHDFFKSPAMPEGYVQERTIFFKDDGNKTKRAEVKFECDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLLGFVTAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_LIKE.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 262BE450E748E44 CRC64;

Query Match 94.3%; Score 1200; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 2e-93;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIILVELDGVNKHGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGELFTGVVPIILVELDGVNKHGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTALSYGVCFSRYPDHMKQHDFFKSPAMPEGYVQERTIFFKDDGNKTKRAEVKFECDTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSPAMPEGYVQERTIFFKDDGNKTKRAEVKFECDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLLGFVTAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
```

```
AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_LIKE.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;

Query Match 93.1%; Score 1185; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 3.8e-92;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIILVELDGVNKHGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGELFTGVVPIILVELDGVNKHGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTALSYGVCFSRYPDHMKQHDFFKSPAMPEGYVQERTIFFKDDGNKTKRAEVKFECDTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSPAMPEGYVQERTIFFKDDGNKTKRAEVKFECDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLLGFVTAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL3916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
```

```
DR PRINTS; PR01229; GFP; FLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 688FD75E88926903 CRC64;

Query Match      84.8%; Score 1080; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 2.8e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKLTIKFTCTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFTCTTGKLPVWPPTL 60

QY 52 VTALSGVOCFSYPDHMKQHDFFKSAMPEGYVQERTIPFKDGNKYKTRAEVKFEGDTLV 121
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 51 VTTLSTGIGCFARYPEHMKNDFFKSAMPEGYIQERTIPFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRLEKGMDFKEDGNILGHKLEYNHSHVYIMPDKANGLKYNFKIRHNIEGGVOLAD 180

QY 132 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNPKRDHMLLGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 HYQTNVPLGDGPVLLPINHYLSQTQTAISKDRNETRDHMLVLEFFSACGHTGMDELYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR0009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.2%; Score 1072; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.3e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKLTIKFTCTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGVVPILVELDGDVHGKFSVRGEGDADYKLEIKFTCTTGKLPVWPPTL 60

QY 62 VTALSGVOCFSYPDHMKQHDFFKSAMPEGYVQERTIPFKDGNKYKTRAEVKFEGDTLV 121
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTLSTGIGCFARYPEHMKNDFFKSAMPEGYIQERTIPFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRLEKGMDFKEDGNILGHKLEYNHSHVYIMPDKANGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNPKRDHMLLGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLLPINHYLSQTQTAISKDRNETRDHMLVLEFFSACGHTGMDELYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR0009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.5%; Score 1076; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 6.1e-83;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKLTIKFTCTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFTCTTGKLPVWPPTL 60

QY 52 VTALSGVOCFSYPDHMKQHDFFKSAMPEGYVQERTIPFKDGNKYKTRAEVKFEGDTLV 121
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 51 VTTLSTGIGCFARYPEHMKNDFFKSAMPEGYIQERTIPFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRLEKGMDFKEDGNILGHKLEYNHSHVYIMPDKANGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNPKRDHMLLGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLLPINHYLSQTQTAISKDRNETRDHMLVLEFFSACGHTGMDELYK 238
```


[illegible]

```

RN  SEQUENCE FROM N.A.
RP  STRAIN:GFPxm19uv;
RC  Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA  Li S.J., Xia N.S.;
RT  "Colorful mutants of green fluorescent protein from Aequorea
RT  macrodactyla.";
RL  Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF435430; AAL33915.1; -.
DR  GO; GO:0006091; P:energy pathways; IEA.
DR  InterPro; IPR009017; GFP like.
DR  InterPro; IPR000786; Green_fl_protein.
DR  Pfam; PF01353; GFP; 1.
DR  PRINTS; PR01229; GFPLORESCENT.
DR  ProDom; PD013756; Green_fl_protein; 1.
SQ  SEQUENCE 238 AA; 27002 MW; 5258A2982264C018 CRC64;

Query Match      83.8%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 3.5e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY  2 VSKGEELFTGVVPILVELDGDVNGHKFSPVSGEGGDAATYKGLTKLEIKTCTGKLPVWPPTL 61
Db  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGIVPELVIELDGDVHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 60

QY  52 VTALSYGVOCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Db  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 VTTLSYGIILCFARYPEHMKMNDFFKSAPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY  122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRNIIDSGVOLAD 181
Db  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 NRIELKGMDFKEDGNILGHKLEYNFNHNYIMPDKANNGLKVNFKIRNIIEGGVOLAD 180

QY  132 HYQONTPIGDGPVLLPDNHYLSQSALSXDPNPKRDHMLLGFVTAAGITLGMDELYK 239
Db  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 HYQTNVPLGDGPVLIPINHYLSQTAISKORNETRDHMLVLEFFSACGTHGMDELYK 238

RESULT 14
Q8WTC5 PRELIMINARY; PRT; 238 AA.
ID Q8WTC5;
AC Q8WTC5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Orange fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OPFxm;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435432; AAL33917.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534E573A CRC64;

Query Match      83.7%; Score 1065; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 5.2e-82;
Matches 195; Conservative 17; Mismatches 26; Indels 0; Gaps 0;

QY  2 VSKGEELFTGVVPILVELDGDVNGHKFSPVSGEGGDAATYKGLTKLEIKTCTGKLPVWPPTL 61

```

```

Db  1 MSKGEELFTGVVPILVELDGDVHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 60
QY  62 VTALSYGVOCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Db  61 VTTLSYGIILCFARYPEHMKMNDFFKSAPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY  122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRNIIDSGVOLAD 181
Db  121 NRIELKGMDFKEDGNILGHKLEYNFNHNYIMPDKANNGLKVNFKIRNIIEGGVOLAD 180
QY  182 HYQONTPIGDGPVLLPDNHYLSQSALSXDPNPKRDHMLLGFVTAAGITLGMDELYK 239
Db  181 HYQTNVPLGDGPVLIPINHYLSQTAISKORNETRDHMLVLEFFSACGTHGMDELYK 238

RESULT 15
Q95UA7 PRELIMINARY; PRT; 225 AA.
ID Q95UA7;
AC Q95UA7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cyan fluorescent protein (Fragment).
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Falkowski P.G., Sun Y.;
RA "Montastraea cavernosa fluorescent protein.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056460; AAL17905.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON TER 225
SQ SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;

Query Match      20.1%; Score 256.5; DB 5; Length 225;
Best Local Similarity 31.6%; Pred. No. 9.9e-14;
Matches 65; Conservative 43; Mismatches 81; Indels 17; Gaps 7;

QY  12 VVPILVELDGDVNGHKFSPVSGEGGDAATYKGLTKLF-ICTTGKLPVWPPTLVALSYGQ 70
Db  7 VMKIKRLMDGIVNGHKFMITGEGGKPFEGTHIILKVKEGGLPFAYDILTTFAYQGNR 66
QY  71 CFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELK 130
Db  67 VFTKYPKDP-DYFKQSPFEGYSWERSMTFEDQGVCTVTSIDIKLEGDCFFYRIFYGVN 124
QY  131 FKEDGNILGHK-LEYNYNHNYIMADKQNGIKVNFKIRNIIDSGVOLADHYQONTPI 189
Db  125 FPSGSPVMQKTKLWEPSTENMYV-----RDGVLGDSVSRLLLEGD----KHRCNFRS 175
QY  190 GDGP---VLLPDNHYLSQ-SALSKD 211
Db  176 TYGAKGVLPPEYHFVDHRIEILSHD 201

Search completed: June 21, 2004, 16:00:08
Job time : 30.7778 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds
(without alignments)
1433.395 Million cell updates/sec

Title: US-09-887-784-64G
Perfect score: 1275
Sequence: 1 MWSKSELTGTGVVLPILVELD.....VLLGFVTAAGITLGMDELK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1267	99.4	239	5 AAE17518	Aae17518 Enhanced
2	1267	99.4	363	6 ABR40352	Abr40352 Human ami
3	1267	99.4	893	4 AAG65781	Aag65781 Amino aci
4	1267	99.4	1132	4 AAG65782	Aag65782 Amino aci
5	1259	98.7	239	5 AAE17517	Aae17517 Enhanced
6	1256	98.5	239	3 AAB22882	Aab22882 Enhanced
7	1256	98.5	239	3 AAY54349	Aay54349 Amino aci
8	1256	98.5	239	3 AAY79584	Aay79584 EGFP sign
9	1256	98.5	239	4 AAB50804	Aab50804 Jellyfish
10	1256	98.5	239	4 AAB85900	Aab85900 A. victor
11	1256	98.5	239	4 AAB31171	Aab31171 Amino aci
12	1256	98.5	239	5 AAG66198	Aag66198 A. victor
13	1256	98.5	239	5 ABG94444	Abg94444 Protease
14	1256	98.5	239	5 AAE14599	Aae14599 Aequorea
15	1256	98.5	239	6 AAE34958	Aae34958 Aequorea
16	1256	98.5	239	6 AAG79829	Aag79829 Green flu
17	1256	98.5	239	6 ABR83616	Abr83616 Green flu
18	1256	98.5	239	6 ADA38074	Ada38074 Aequorea
19	1256	98.5	239	7 ABUC13204	Abuc13204 Aequorea
20	1256	98.5	239	7 ADC18358	Adc18358 EGFP (enh
21	1256	98.5	239	7 ABW00914	Abw00914 Aequorea
22	1256	98.5	239	7 ADE28570	Ade28570 Enhanced
23	1256	98.5	246	7 ABM79011	Abm79011 Enhanced
24	1256	98.5	248	5 AAG68319	Aag68319 Jellyfish
25	1256	98.5	259	5 AAU99804	Aau99804 Biomembra

26	1256	98.5	265	2 AAW97451	Aaw97451 Wild-type
27	1256	98.5	268	5 AAU99803	Aau99803 Biomembra
28	1256	98.5	270	5 AAU99802	Aau99802 Biomembra
29	1256	98.5	272	5 AAU99800	Aau99800 Biomembra
30	1256	98.5	273	5 AAU99801	Aau99801 Biomembra
31	1256	98.5	280	5 AAU99807	Aau99807 Biomembra
32	1256	98.5	281	3 AAY50142	Aay50142 Green flu
33	1256	98.5	281	3 AAB24252	Aab24252 EGFP-MODC
34	1256	98.5	281	5 AAU10888	Aau10888 EGFP-MODC
35	1256	98.5	286	7 ADE28562	Ade28562 EGFP/ hum
36	1256	98.5	289	7 ADE28564	Ade28564 EGFP/ hum
37	1256	98.5	290	7 ADE28568	Ade28568 EGFP/ hum
38	1256	98.5	290	7 ADE28566	Ade28566 EGFP/ hum
39	1256	98.5	294	3 AAB22860	Aab22860 GFP-DEVD-
40	1256	98.5	294	3 AAY79638	Aay79638 Caspase-3
41	1256	98.5	294	5 ABG94422	Abg94422 Recombina
42	1256	98.5	308	2 AAY42181	Aay42181 EGFP/DRM
43	1256	98.5	320	6 ABR83620	Abr83620 HUB1-GFP
44	1256	98.5	323	3 AAY54359	Aay54359 GFP mutan
45	1256	98.5	323	6 ABR83621	Abr83621 RUB1-GFP

ALIGNMENTS

RESULT 1	
AAE17518	
ID	AAE17518 standard; protein; 239 AA.
XX	XX
AC	AAE17518;
XX	XX
DT	22-APR-2002 (first entry)
XX	XX
DE	Enhanced F64L-E222G jellyfish green fluorescent protein mutant.
XX	XX
XW	Jellyfish; green fluorescent protein; GFP; protein redistribution;
XW	cellular function; genetic reporter; mutant; Stoke's shift; mutein.
OS	Aequorea victoria.
OS	Synthetic.
XX	XX
FH	Key Location/Qualifiers
FT	Misc-difference 65
FT	/note= "Wild type Phe substituted with Leu; This corresponds to position 64 in the wild type protein"
FT	Misc-difference 223
FT	/note= "Wild type Glu substituted with Gly; This corresponds to position 222 in the wild type protein"
XX	XX
PN	WO200198338-A2.
XX	XX
PD	27-DEC-2001.
XX	XX
PF	18-JUN-2001; 2001WO-EP006848.
XX	XX
PR	19-JUN-2000; 2000DK-00000953.
PR	20-JUN-2000; 2000US-0212681P.
PR	10-MAY-2001; 2001DK-00000739.
PR	10-MAY-2001; 2001US-0290170P.
XX	XX
PA	(BIOI-) BIOIMAGE AS.
XX	XX
PI	Bjorn SP, Pagliaro L, Thastrup O;
XX	XX
DR	WPI; 2002-098224/13.
XX	XX
DR	N-PSDB; AAD28163.
XX	XX
PT	Novel fluorescent protein in in vitro assay for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution, has a green fluorescent protein with F64L and E222G mutation.
XX	XX
PS	Claim 9; Page 37; 41pp; English.

XX The invention relates to a fluorescent protein derived from green
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
CC at F64L and E222G has a bigger compared to other GFP's making it very
CC suitable for high throughput screening due to better resolution. The
CC fluorescent protein is useful in vitro assays for measuring protein
CC kinase activity or dephosphorylation activity, or for measuring protein
CC redistribution. The fluorescent protein is useful in studying cellular
CC functions in living cells; as protein tags in transgenic animals, living
CC and fixed cells; organelle tags, secretion marker and genetic reporter.
CC The fluorescent protein is also useful as a cell or organelle integrity
CC marker, a marker for changes in cell morphology, as transfection marker,
CC and as a marker to be used in combination with fluorescence activated
CC cell sorting (FACS). The novel proteins can also be used as reporters to
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
CC protein is also useful as markers in transcriptional and translational
CC fusions for performing transposon vector mutagenesis and as a reporter
CC for bacterial detection. Transposons encoding the fluorescent protein are
CC useful for screening promoters and for tagging plasmids and chromosomes.
CC The fluorescent protein engineered into the genome of a phage is useful
CC for designing diagnostic tool. The present sequence is a DNA encoding
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
XX
SQ Sequence 239 AA;

Query Match 99.4%; Score 1267; DB 5; Length 239;
Best Local Similarity 99.8%; Pred. No. 1.3e-122;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTKLFICTTGKLPVPWPT 60

QY 51 LVTGLSYGVQCFSRYPDHNKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 51 LVTGLSYGVQCFSRYPDHNKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 2
ABR40352
ID ABR40352 standard; protein; 363 AA.
XX
AC ABR40352;
XX
DT 08-JUL-2003 (first entry)
DE Human amino acid sequence SEQ ID NO: 6.
DE Human; heterologous conjugate; intracellular protein.
XX
OS Homo sapiens.
OS Aequoria victoria.
XX
PN WO2003029827-A2.
XX
PD 10-APR-2003.
XX
PF 01-OCT-2002; 2002WO-DK000651.
XX
PR 01-OCT-2001; 2001DK-00001433.
XX
PR 11-OCT-2001; 2001US-032896P.
XX
PA (BIOL-) BIOIMAGE AS.
XX
PI Terry BR, Nielsen SJ;

XX WPI; 2003-430211/40.
DR N-PSDB; ACC72604.
XX
PT Novel cell for identifying modulators of protein interaction, contains a
PT first conjugate comprising anchor protein, second conjugate having type B
PT interactor protein and third conjugate with detectable group.
XX
PS Disclosure; Page 112-113; 118pp; English.
XX
CC The invention relates to a novel cell, comprising three heterologous
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that
CC specifically binds to an internal structure within the cell conjugated to
CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
CC type B conjugated to a first protein of interest, and a third HC (HC3)
CC comprising a second protein of interest conjugated to detectable group.
CC The cell is useful for detecting if a compound disrupts or induces the
CC interaction between two intracellular proteins. The cell is also useful
CC for screening compounds that modulate the interaction between two
CC intracellular proteins. The present sequence is used in the
CC exemplification of the invention
XX
SQ Sequence 363 AA;

Query Match 99.4%; Score 1267; DB 6; Length 363;
Best Local Similarity 99.6%; Pred. No. 2.4e-122;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTGLSYGVQCFSRYPDHNKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTGLSYGVQCFSRYPDHNKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3
AAG65781
ID AAG65781 standard; protein; 893 AA.
XX
AC AAG65781;
XX
DT 07-JAN-2002 (first entry)
DE Amino acid sequence of HSPDE4A1-E222G fusion protein.
DE PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX fusion protein.
XX
OS Homo sapiens.
OS Aequorea victoria.
XX
PN WO200179526-A2.
XX
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-DK000264.
XX
PR 17-APR-2000; 2000DK-00000651.
XX
PR 29-MAY-2000; 2000DK-00000849.
XX
PA (BIOI-) BIOIMAGE AS.
XX

```
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66852.
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX Example 1; Page 156-160; 160pp; English.
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
XX protein
XX Sequence 893 AA;
XX Query Match 99.4%; Score 1267; DB 4; Length 893;
XX Best Local Similarity 99.6%; Pred. No. 9.1e-122;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEBELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPFICTTGKLPVPWPT 60
Db 655 MVSKEBELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPFICTTGKLPVPWPT 714
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 715 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 775 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 835 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 893
RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA.
XX AAG65782;
AC AAG65782;
XX 07-JAN-2002 (first entry)
DT 07-JAN-2002 (first entry)
XX Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX fusion protein.
XX Homo sapiens.
XX Aequorea victoria.
XX OS
XX WO200179526-A2.
XX PN
XX DE
XX
```

```
PD 25-OCT-2001.
XX 11-APR-2001; 2001WO-DK000264.
XX 17-APR-2000; 2000DK-00000651.
PR 29-MAY-2000; 2000DK-00000849.
XX (BIOI-) BIOIMAGE AS.
XX Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
XX Praestegaard M;
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66853.
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX Example 1; Page 162-167; 160pp; English.
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
XX protein
XX Sequence 1132 AA;
XX Query Match 99.4%; Score 1267; DB 4; Length 1132;
XX Best Local Similarity 99.6%; Pred. No. 1.3e-121;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEBELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPFICTTGKLPVPWPT 60
Db 894 MVSKEBELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPFICTTGKLPVPWPT 953
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 954 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 1014 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 1074 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 1132
RESULT 5
AAEI7517
ID AAEI7517 standard; protein; 239 AA.
XX AAEI7517;
AC AAEI7517;
XX 22-APR-2002 (first entry)
DT 22-APR-2002 (first entry)
XX Enhanced F64L jellyfish green fluorescent protein mutant.
XX DE
XX
```

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
 KW cellular function; genetic reporter; mutant; Stoke's shift; mutain.
 XX
 OS Aequorea victoria.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 65 /note= "wild type Phe substituted with Leu; This
 FT corresponds to position 64 in the wild type protein"
 FT
 PN WO200198338-A2.
 XX
 XX 27-DEC-2001.
 XX
 XX 19-JUN-2001; 2001WO-EP006848.
 XX
 XX 19-JUN-2000; 2000DK-00000953.
 PR 20-JUN-2000; 2000US-0212681P.
 PR 10-MAY-2001; 2001DK-00000739.
 PR 10-MAY-2001; 2001US-0290170P.
 XX
 XX (BIOI-) BIOIMAGE AS.
 PA
 XX Bjorn SP, Pagliaro L, Thastrup O;
 XX
 XX WPI; 2002-098224/13.
 DR N-PSDB; AAD28162.
 XX
 XX Novel fluorescent protein in in vitro assay for measuring protein kinase
 PT activity or dephosphorylation activity, or for measuring protein
 PT redistribution, has a green fluorescent protein with F64L and E222G
 PT mutation.
 FT
 XX
 PS Example 1; Page 35; 41pp; English.
 XX
 XX The invention relates to a fluorescent protein derived from green
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
 CC at F64L and E222G has a bigger compared to other GFP's making it very
 CC suitable for high throughput screening due to better resolution. The
 CC fluorescent protein is useful in invitro assays for measuring protein
 CC kinase activity or dephosphorylation activity, or for measuring protein
 CC redistribution. The fluorescent protein is useful in studying cellular
 CC functions in living cells; as protein tags in transgenic animals, living
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.
 CC The fluorescent protein is also useful as a cell or organelle integrity
 CC marker, a marker for changes in cell morphology, as transfection activated
 CC and as a marker to be used in combination with fluorescence activated
 CC cell sorting (FACS). The novel proteins can also be used as reporters to
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
 CC protein is also useful as markers in transcriptional and translational
 CC fusions for performing transposon vector mutagenesis and as a reporter
 CC for bacterial detection. Transposons encoding the fluorescent protein are
 CC useful for screening promoters and for tagging plasmids and chromosomes.
 CC The fluorescent protein engineered into the genome of a phage is useful
 CC for designing diagnostic tool. The present sequence is enhanced F64L
 CC Jellyfish green fluorescent protein (GFP) mutant
 XX
 SQ Sequence 239 AA;
 Query Match 98.7%; Score 1259; DB 5; Length 239;
 Best Local Similarity 99.2%; Pred. No. 8,7e-122;
 Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVPIILVELDGVNKHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
 DB 1 MVSKEELFTGVPIILVELDGVNKHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
 QY 61 LVTLGSLGVOCFGRYPDPMKHQDFFKSAPEGVYQERTIFFKDDGNVYKTAEVKFGDTL 120
 DB 61 LVTTLSGVQCFGRYPDPMKHQDFFKSAPEGVYQERTIFFKDDGNVYKTAEVKFGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 RESULT 6
 AAB22882
 ID AAB22882 standard; protein; 239 AA.
 XX
 AC AAB22882;
 XX
 DT 10-JAN-2001 (first entry)
 XX
 DE Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.
 XX
 KW Bioreporter protein; fusion protein; recognition site;
 KW cellular targeting sequence; cellular localisation; fluorescent protein;
 KW protease activity detection; toxin detection; cellular stress detection;
 KW drug discovery; cell based screening.
 XX
 OS Aequorea victoria.
 OS Synthetic.
 XX
 PN WO200050872-A2.
 XX
 PD 31-AUG-2000.
 XX
 XX 25-FEB-2000; 2000WO-US004794.
 PF
 XX 26-FEB-1999; 99US-0122152P.
 PR 08-MAR-1999; 99US-0123399P.
 PR 12-JUL-1999; 99US-00352171.
 XX
 XX (CELL-) CELLOMICS INC.
 PA
 XX Giuliano KA, Kapur R;
 PI
 XX WPI; 2000-594086/56.
 DR N-PSDB; AAA93373.
 XX
 XX Automated cell-based characterization of toxin by contacting cells
 PT containing luminescent reporter molecules with test substance and
 PT analyzing optically.
 XX
 PS Example 11; Fig 29A; 336pp; English.
 XX
 XX The invention relates to systems, methods and reagents for cell-based
 CC screening or detection of compounds which affect particular biological
 CC functions. The methods of the invention utilise fluorescent bioreporter
 CC molecules which, when acted on by a compound of interest, cause an
 CC alteration in the cellular distribution of at least the fluorescent
 CC moiety. In one embodiment, the biosensors comprise heat shock proteins
 CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent
 CC protein (GFP), or derivatives thereof). Such biosensors are located in
 CC the cytoplasm, but on stress activation translocate to the nucleus. In
 CC another embodiment bioreporter proteins can be used to detect protease
 CC activity. Such protease bioreporter fusion proteins comprise one or more
 CC fluorescent proteins; a recognition signal which is cleaved by the
 CC protease; and at least one cellular localisation signal. The latter two
 CC components may be components of a single protein which is acted upon by
 CC the protease, or may be from heterologous sources. Due to the
 CC localisation signal, the bioreporter protein is localised to a particular
 CC region of the cell. Once acted on by the protease of interest, the
 CC fluorescent protein is cleaved from the localisation sequence, and is
 CC free to migrate to other locations within the cell. The presence of a
 CC second localisation signal attached to the fluorescent protein enables
 CC the fluorescent protein to be directed to a different cellular
 CC compartment after cleavage of the protease recognition sequence. The
 CC change in distribution of the fluorescent protein can be detected using
 CC imaging methods with a high degree of spatial resolution. The methods and

CC biosensors of the invention can be used to investigate a wide range of
CC cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased quantity
CC and speed of data collection and faster evaluation of drug candidates.
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used
CC as components of biosensor fusion proteins of the invention
XX
SQ Sequence 239 AA;

Query Match 98.5%; Score 1256; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.8e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPKFICTTGKLPVWPWT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 7
AAV54349
ID AAY54349 standard; protein; 239 AA.
AC AAY54349;
DT 06-APR-2000 (first entry)
DE Amino acid sequence of the mutant green fluorescent protein EGFP.
XX Fluorescent protein; green fluorescent protein; emission intensity;
KW fluorescence; pH detection; pH sensor; EGFP.
XX Synthetic.
OS Aequorea victoria.

XX Key Location/Qualifiers
FH Misc-difference 65
FT /note= "wild type Phe substituted with Leu"
FT Misc-difference 66
FT /note= "wild type Ser substituted with Thr"
FT Misc-difference 232
FT /note= "wild type His substituted with Leu"
XX
PN WO9964592-A2.
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US012850.
XX
PR 09-JUN-1998; 98US-00094359.
PR 13-OCT-1998; 98US-00172063.
XX
XX (REG) UNIV CALIFORNIA.
PA (UYOR-) UNIV OREGON STATE.
XX

PI Tsien RY, Llopis J, Wachter RM;
XX WPI; 2000-116540/10.
DR N-PSDB; AA245642.
XX
XX New functional engineered green fluorescent proteins, used for measuring
XX the pH in biological samples and cells.
XX
XX Disclosure; Page 9; 89pp; English.
CC The present sequence represents a functional engineered fluorescent
CC protein based on the Aequorea green fluorescent protein (GFP). The
CC emission intensity changes as pH varies between 5 and 10 of the present
CC protein are novel. The functional engineered fluorescent proteins show
CC reversible changes in fluorescence over physiological pH ranges. They can
CC be used for determining the pH of samples and cells. The polynucleotides
CC can also be used to produce transgenic animals. The fluorescent protein
CC pH sensors can be delivered to cells in the form of polynucleotides
CC encoding the protein sensor fused to a targeting signal. The targeting
CC signal directs the expression of the protein sensors to restricted cell
CC locations. This makes it possible to measure the pH of a precisely
CC defined cellular region or organelle
XX
SQ Sequence 239 AA;

Query Match 98.5%; Score 1256; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.8e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPKFICTTGKLPVWPWT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 8
AAV79584
ID AAY79584 standard; peptide; 239 AA.
XX
AC AAY79584;
DT 29-AUG-2000 (first entry)
XX
DE EGFP signal domain.
XX
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;
XX analysis; drug discovery.
XX Unidentified.
XX
PN WO200026408-A2.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99WO-US025431.
XX
PR 30-OCT-1998; 98US-0106308P.
PR 26-MAY-1999; 99US-0136078P.
XX
XX (CELL-) CELLOMICS INC.
XX
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX WPI; 2000-365644/31.
 DR N-PSDB; AAA27573.
 XX
 PT Recombinant nucleic acid encoding a protease biosensor useful for
 PT fluorescence based cell and molecular biochemical assays for drug
 PT discovery comprising three operably linked nucleic acid sequences.
 XX
 XX
 PS Claim 14; Fig 29A; 218pp; English.
 XX
 CC The present sequence is that of the EGFP signal domain, which can be
 CC included in novel recombinant protease biosensors (PBs) of the invention.
 CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)
 CC comprising at least 1 detectable polypeptide signal such as the present
 CC sequence; a second domain (see AAY79588-622) comprising at least 1
 CC protease recognition site; and a third domain (see AAY79623-37)
 CC comprising at least 1 reactant target sequence. A recombinant nucleic
 CC acid (see AAA27627-43) encoding the PB, an expression vector, and a
 CC genetically engineered host cell are also claimed. A claimed method for
 CC identifying compounds that modify protease activity in a cell involves
 CC contacting a host cell that possesses the recombinant PB with a test
 CC compound, and determining the PB distribution in the host cell, where
 CC changes in the distribution of the PB are correlated with modification of
 CC protease activity by the test compound. Claimed kits for identifying
 CC compounds that modify protease activity in a host cell include the
 CC recombinant nucleic acid, or the recombinant PB, or the vector, or the
 CC host cell. The PB is useful in high content screens to detect in vivo
 CC activation of enzymatic activity, and to identify specific activity based
 CC on cleavage of a known recognition motif
 XX
 SQ Sequence 239 AA;
 Query Match 98.5%; Score 1256; DB 3; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.8e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
 QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDIL 120
 DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDIL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
 RESULT 9
 AAB50804
 ID AAB50804 standard; protein; 239 AA.
 XX
 AC AAB50804;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Jellyfish GFP mutant EGFP.
 XX
 KW Aequorea victoria; jellyfish; fluorescent protein indicator;
 KW green fluorescent protein; GFP; linker moiety; sensor;
 KW calmodulin-binding domain; mutant; muten.
 XX
 OS Aequorea victoria.
 XX
 PN WO200071565-A2.
 XX
 PD 30-NOV-2000.
 XX

PF 17-MAY-2000; 2000WO-US013684.
 XX
 PR 21-MAY-1999; 99US-00316919.
 PR 21-MAY-1999; 99US-00316920.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Tsien RY, Baird GA;
 XX
 DR WPI; 2001-032017/04.
 DR N-PSDB; AAC90488.
 XX
 PT Novel fluorescent proteins comprising a sensor protein inserted into
 PT them, useful for measuring the response of a sensor biological, chemical,
 PT electrical or physiological parameter in vivo or in vitro.
 XX
 PS Disclosure; Page 24; 94pp; English.
 XX
 CC The present sequence is a fluorescent protein used in the construction of
 CC a fluorescent protein indicator. The indicator comprises a sensor
 CC polypeptide that is responsive to a chemical, biological, electrical or
 CC physiological parameter, and a fluorescence protein functional group. The
 CC sensor polypeptide is operatively inserted into the fluorescent moiety.
 CC The fluorescent indicator is useful for detecting the presence of a
 CC response inducing member in a sample. The method involves contacting the
 CC sample with the indicator and detecting a change in fluorescence, in
 CC which a change is indicative of the effect of the parameter on the sensor
 CC polypeptide. The novel fluorescent proteins are advantageous due to their
 CC reduced size as compared to the FRET (fluorescence resonance energy
 CC transfer)-based sensors
 XX
 SQ Sequence 239 AA;
 Query Match 98.5%; Score 1256; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.8e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
 QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDIL 120
 DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDIL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
 RESULT 10
 AAB85900
 ID AAB85900 standard; protein; 239 AA.
 XX
 AC AAB85900;
 XX
 DT 30-NOV-2001 (first entry)
 XX
 DE A. victoria green fluorescent protein (GFP) and linker sequence.
 XX
 KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
 KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;
 KW green fluorescent protein.
 XX
 OS Synthetic.
 OS Aequorea victoria.
 XX
 PN WO200168706-A1.
 XX

PD 20-SEP-2001.
 XX
 PF 14-MAR-2001; 2001WO-US008071.
 XX
 PR 15-MAR-2000; 2000US-0189698P.
 XX
 XX (MERI) MERCK & CO INC.
 PA
 XX Marsh DJ;
 XX
 XX WPI; 2001-565791/63.
 DR N-PSDB; AAH47304.
 XX
 XX Fusion proteins comprising melanin concentrating hormone receptor
 PT peptides and fluorescent proteins, useful for identifying appetite
 PT stimulants.
 XX
 XX Claim 2; Page 14; 71pp; English.
 XX
 XX The invention provides melanin concentrating hormone (MCH) receptor
 CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
 CC MCHR polypeptide regions from different species. The MCHR fusion protein
 CC comprise MCHR polypeptide region and a fluorescent polypeptide region
 CC joined directly, or via a linker, to the carboxy side of the MCHR
 CC polypeptide region. The MCHR fusion proteins can be expressed by standard
 CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
 CC regulation of MCH activity stimulates food intake. The present sequence
 CC represents a A. victoria green fluorescent protein (GFP) and a linker
 CC sequence
 XX
 XX Sequence 239 AA;
 SQ
 Query Match 98.5%; Score 1256; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.8e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLFICTTGTGLPVPWPT 60
 Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLFICTTGTGLPVPWPT 60
 QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 Db 61 LVTTLTYGQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
 Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
 QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 Db 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 RESULT 11
 AAB31171
 ID AAB31171 standard; protein; 239 AA.
 XX
 AC AAB31171;
 XX
 XX 02-APR-2001 (first entry)
 DT
 XX Amino acid sequence of a green fluorescent protein (GFP).
 DE
 XX Growth rate; death rate; reporter gene; luminescent protein;
 KW fluorescent product; luciferase; green fluorescent protein; GFP.
 KW
 XX Aequorea victoria.
 OS
 XX WO200075367-A1.
 PN
 XX 14-DEC-2000.
 PD
 XX 07-JUN-2000; 2000WO-FI000507.
 PF

XX
 PR 07-JUN-1999; 99FI-00001296.
 XX
 PA (LILI/) LILIUS E.
 PA (VIRT/) VIRT M.
 XX
 PI Lilius E, Virta M;
 XX
 XX WPI; 2001-061737/07.
 DR N-PSDB; AAC86954.
 DR
 XX
 XX Assessing growth and death rates of a micro-organism in a desired
 PT environment, by introducing 2 reporter genes encoding luminescent and
 PT fluorescent products and detecting luminescent fluorescence.
 XX
 XX Disclosure; Page 27; 32pp; English.
 XX
 XX The specification describes a method for assessing the growth rate and
 CC death rate of a micro-organism within a predetermined time period in a
 CC desired environment. The method comprises introducing at least two
 CC reporter genes encoding luminescent and/or fluorescent products into the
 CC micro-organisms, incubating the micro-organism within the desired
 CC environment, and detecting luminescence and/or fluorescence after a
 CC predetermined time period. Use of two different markers within a micro-
 CC organism enables the differentiation between growth and death rates. The
 CC method is used to assess the growth rate and death rate of a micro-
 CC organism within a predetermined time period in a desired environment. The
 CC present sequence represents a green fluorescent protein (GFP), and is
 CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
 CC and is used in the method of the invention
 XX
 XX Sequence 239 AA;
 SQ
 Query Match 98.5%; Score 1256; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.8e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLFICTTGTGLPVPWPT 60
 Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLFICTTGTGLPVPWPT 60
 QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 Db 61 LVTTLTYGQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
 Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
 QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 Db 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 RESULT 12
 AAG66198
 ID AAG66198 standard; protein; 239 AA.
 XX
 AC AAG66198;
 XX
 XX 17-JUN-2002 (first entry)
 DT
 XX A. victoria green fluorescent protein (EGFP).
 DE
 XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
 KW green fluorescent protein; EGFP.
 KW
 XX Aequorea victoria.
 OS
 XX JP2002045189-A.
 PN
 XX 12-FEB-2002.
 PD
 XX

```
PF 04-AUG-2000; 2000JP-00237165.
XX
PR 04-AUG-2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2002-299190/34.
XX N-PSDB; ABL40628.
XX
XX A gene encoding cyan-green fluorescent protein.
XX
XX Example; Page 14; 20pp; Japanese.
XX
XX The invention relates to a gene encoding proteins having cyan-green
XX fluorescence characteristic and having a function of showing stable
XX fluorescence characteristic in acid region. A method for the preparation
XX of a cyan-green fluorescent protein is provided which involves a
XX transformant transformed by a recombinant vector comprising the gene,
XX where the transformant is cultured and the protein is collected from the
XX culture. The present sequence represents the A. victoria green
XX fluorescent protein (EGFP)
XX
XX Sequence 239 AA;
XX
XX Query Match 98.5%; Score 1256; DB 5; Length 239;
XX Best Local Similarity 98.7%; Pred. No. 1.8e-121;
XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 13
ABG94444
ID ABG94444 standard; protein; 239 AA.
XX
XX AC ABG94444;
XX
XX DT 27-NOV-2002 (first entry)
XX
XX Protease biosensor signal sequence #6.
XX
XX DE Detection; classification; identification; toxin detection; protease;
XX KW ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
XX KW toxic threat agent.
XX
XX OS Synthetic.
XX
XX US6416959-B1.
XX
XX PD 09-JUL-2002.
XX
XX PF 25-FEB-2000; 2000US-00513783.
XX
XX PR 27-FEB-1997; 97US-00810983.
XX PR 27-FEB-1998; 98US-00031271.
XX PR 26-FEB-1999; 99US-0122152P.
XX PR 08-MAR-1999; 99US-0123399P.
XX PR 12-JUL-1999; 99US-00352171.
XX PR 31-AUG-1999; 99US-0151797P.
XX
PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
XX 01-DEC-1999; 99US-0168408P.
XX
XX (GIUL/) GIULIANO K.
XX PA (KAFU/) KAFUR R.
XX
XX Giuliano K, Kapur R;
XX
XX WPI; 2002-634730/68.
XX N-PSDB; ABS71491.
XX
XX Automated cell-based toxin detection, classification, and/or
XX identification by treating cells involves use of three classes of
XX luminescent reporter molecules such as detectors, classifiers or
XX identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
XX
XX The invention describes methods of automated detection, classification
XX and identification comprising treating cells containing luminescent
XX reporter molecules (I) in array of locations with a test substance, where
XX (I) are detectors, classifiers or identifiers, imaging cells in each
XX location to obtain luminescent signals and converting optical information
XX into digital data to interpret presence of toxins in the test substance.
XX The method are useful for detection of toxins chosen from proteases, ADP-
XX ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
XX Three classes of cell-based luminescent reporter molecules such as
XX detectors, classifiers and identifiers are described and serve as
XX reporters of toxic threat agents. The first two levels of
XX characterization enable a rapid readout of toxin class without
XX sacrificing the ability to detect many new mutant toxins or dissect
XX several complex mixtures of known toxins. This is the amino acid sequence
XX of a protease biosensor related signal sequence used in the cell-based
XX screening system
XX
XX SQ Sequence 239 AA;
XX
XX Query Match 98.5%; Score 1256; DB 5; Length 239;
XX Best Local Similarity 98.7%; Pred. No. 1.8e-121;
XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AC AAE14599;
XX
XX DT 31-MAY-2002 (first entry)
XX
XX DE Aequorea victoria enhanced green fluorescent protein.
XX KW Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; mutein.
XX OS Aequorea victoria.
XX OS Synthetic.
```

```

XX Key Location/Qualifiers
PH Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"
FT
FT
FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"
FT
FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"
FT
FT
XX EPl178109-A1.
XX
XX 06-FEB-2002.
XX
XX 03-AUG-2001; 2001EP-00306650.
XX
XX 04-AUG-2000; 2000JP-00237166.
XX
XX (RIKE ) RIKEN KK.
XX
XX Miyawaki A, Sawano A;
XX
XX WPI: 2002-208112/27.
XX N-PSDB; AAD27910.
XX
XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimeres.
XX
XX Example 1; Page 13-14; 31pp; English.
XX
XX The invention relates to a method for mutagenesis that comprises synthesising a mutated strand and a complementary strand by use of megaprimeres. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated primers are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimeres, the megaprimeres are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria
XX
XX Sequence 239 AA;
XX
XX Query Match
XX Best Local Similarity 98.5%; Score 1256; DB 5; Length 239;
XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
XX Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
XX
XX QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDYL 120
XX Db 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDYL 120
XX
XX QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
XX Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
XX
XX QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
XX Db 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239

```

```

QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
Db 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239

RESULT 15
AAE34958
ID AAE34958 standard; protein; 239 AA.
XX
XX AAE34958;
XX
XX 28-MAY-2003 (first entry)
XX
XX Aequorea victoria enhanced green fluorescent protein (EGFP).
XX
XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;
XX kinase; enhanced green fluorescent protein; EGFP.
XX
XX Aequorea victoria.
XX
XX WO200295058-A2.
XX
XX 28-NOV-2002.
XX
XX 24-MAY-2002; 2002WO-US016955.
XX
XX 24-MAY-2001; 2001US-00865291.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Tsien RY, Ting AY, Zhang J;
XX
XX WPI: 2003-148474/14.
XX N-PSDB; AA553428.
XX
XX Novel chimeric phosphorylation indicators, useful for detecting
XX kinase/phosphatase in samples, has donor molecule, phosphorylatable
XX domain, phosphoaminoacid binding domain, and acceptor molecule, in
XX operative linkage.
XX
XX Disclosure; Col 56-57; 38pp; English.
XX
XX The present invention relates to chimeric phosphorylation indicators
XX comprising a phosphorylation polypeptide and a fluorescent protein or in
XX operative linkage, a donor molecule, a phosphorylatable domain, a
XX phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The
XX phosphorylation indicators of the invention are useful for detecting in
XX kinases or phosphatases in a biological sample. They are also useful in
XX high throughput analysis e.g. for detecting a kinase inhibitor or
XX phosphatase inhibitor. The present sequence is Aequorea victoria enhanced
XX green fluorescent protein (EGFP) used in the invention
XX
XX Sequence 239 AA;
XX
XX Query Match
XX Best Local Similarity 98.5%; Score 1256; DB 6; Length 239;
XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
XX Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
XX
XX QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDYL 120
XX Db 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDYL 120
XX
XX QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
XX Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
XX
XX QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
XX Db 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239

```

Db 131 DHYQONTPIGDGPVLLPDNHYLSQSALS KDPNEKEDHNVLLFEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:08
Job time : 50.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-64G
Perfect score: 1275
Sequence: 1 MWSKGBELFTGVVPIVLVD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patente AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pdp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pdp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pdp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pdp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pdp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1256	98.5	239	3	US-09-172-063-3
2	1256	98.5	239	4	US-09-513-783A-46
3	1256	98.5	239	4	US-09-316-919-4
4	1256	98.5	239	4	US-09-602-641-3
5	1256	98.5	239	4	US-09-920-922-2
6	1256	98.5	281	3	US-09-062-102-1
7	1256	98.5	281	4	US-09-364-946-1
8	1256	98.5	294	4	US-09-513-783A-2
9	1256	98.5	323	3	US-09-172-063-21
10	1256	98.5	323	4	US-09-602-641-21
11	1256	98.5	364	3	US-09-085-305-6
12	1256	98.5	379	4	US-09-417-197-129
13	1256	98.5	434	4	US-09-800-170-48
14	1256	98.5	442	4	US-09-417-197-127
15	1256	98.5	459	4	US-09-513-783A-170
16	1256	98.5	544	4	US-09-417-197-113
17	1256	98.5	544	4	US-09-417-197-115
18	1256	98.5	604	4	US-09-417-197-59
19	1256	98.5	605	4	US-09-417-197-41
20	1256	98.5	606	4	US-09-417-197-65
21	1256	98.5	607	4	US-09-417-197-47
22	1256	98.5	630	4	US-09-417-197-63
23	1256	98.5	631	4	US-09-417-197-39
24	1256	98.5	633	4	US-09-417-197-45
25	1256	98.5	635	4	US-09-417-197-125
26	1256	98.5	642	2	US-08-818-252-2
27	1256	98.5	642	2	US-08-818-253-6

28	1256	98.5	642	3	US-08-818-252-2	Sequence 2, Appli
29	1256	98.5	642	3	US-08-818-252-6	Sequence 6, Appli
30	1256	98.5	652	2	US-08-818-253-4	Sequence 4, Appli
31	1256	98.5	652	3	US-08-818-252-4	Sequence 4, Appli
32	1256	98.5	718	4	US-09-417-197-75	Sequence 75, Appli
33	1256	98.5	719	4	US-09-417-197-51	Sequence 51, Appli
34	1256	98.5	726	4	US-09-417-197-71	Sequence 71, Appli
35	1256	98.5	727	4	US-09-417-197-139	Sequence 139, App
36	1256	98.5	783	4	US-09-513-783A-176	Sequence 176, App
37	1256	98.5	797	4	US-09-417-197-141	Sequence 141, App
38	1256	98.5	797	4	US-09-417-197-143	Sequence 143, App
39	1256	98.5	798	4	US-09-417-197-77	Sequence 77, Appli
40	1256	98.5	805	4	US-09-513-783A-178	Sequence 178, App
41	1256	98.5	806	4	US-09-417-197-53	Sequence 53, Appli
42	1256	98.5	836	4	US-09-417-197-61	Sequence 61, Appli
43	1256	98.5	842	4	US-09-417-197-43	Sequence 43, Appli
44	1256	98.5	843	4	US-09-417-197-117	Sequence 117, App
45	1256	98.5	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1
US-09-172-063-3
; Sequence 3, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Liopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-172-063-3

Query Match	98.5%	Score 1256;	DB 3;	Length 239;
Best Local Similarity	98.7%	Pred. No. 1.1e-127;		
Matches 236;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	1	MWSKGBELFTGVVPIVLVELDGDVNGHKFVS	SGEGDATYGKLTLPKICTTGKLPVPWPPT	60
Db	1	MWSKGBELFTGVVPIVLVELDGDVNGHKFVS	SGEGDATYGKLTLPKICTTGKLPVPWPPT	60
Qy	61	LVTGLSYGVQCFSRYPDPHMKQHDFFKSAM	PEGVYQERTIFFKDDGNGYKTRAEVKFEGD	120
Db	61	LVTTLTYGVQCFSRYPDPHMKQHDFFKSAM	PEGVYQERTIFFKDDGNGYKTRAEVKFEGD	120
Qy	121	VNRIELKGDIFDKEDGNILGHKLEYNNY	NSHNVYIMADKQNGIKVNFKIRNIEDGS	180
Db	121	VNRIELKGDIFDKEDGNILGHKLEYNNY	NSHNVYIMADKQNGIKVNFKIRNIEDGS	180
Qy	181	DHYQQNTPIGDGPVLLPDNNHYLSTQ	SALSKDPNEKRDHVMVLLGFVTAAGITL	239
Db	181	DHYQQNTPIGDGPVLLPDNNHYLSTQ	SALSKDPNEKRDHVMVLLGFVTAAGITL	239

```
RESULT 2
US-09-513-793A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kaput, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          98.5%; Score 1256; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLKFICTTGTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLKFICTTGTGKLPVWPWT 60

Qy 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

Qy 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180

Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.5%; Score 1256; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLKFICTTGTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLKFICTTGTGKLPVWPWT 60

Qy 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

Qy 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180

Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wächter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.5%; Score 1256; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLKFICTTGTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLKFICTTGTGKLPVWPWT 60

Qy 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

Qy 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180

Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; 
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match      98.5%; Score 1256; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.4e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
DB 121 VNRIELKGIDPFEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNLSQTSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNLSQTSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; EARLIER FILING DATE: 1998-04-17
; CURRENT FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
US-09-062-102-1

Query Match      98.5%; Score 1256; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.4e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
DB 121 VNRIELKGIDPFEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNLSQTSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNLSQTSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match      98.5%; Score 1256; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.4e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
DB 121 VNRIELKGIDPFEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNLSQTSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNLSQTSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
```

```

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
US-09-364-946-1

Query Match      98.5%; Score 1256; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.4e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
DB 121 VNRIELKGIDPFEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNLSQTSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNLSQTSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match      98.5%; Score 1256; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
|||||
```

```
Db 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60
QY 51 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKEGDTL 120
Db 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKEGDTL 120
QY 121 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9
US-09-172-063-21
; Sequence 21, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; EARLIER APPLICATION NUMBER: 09/094,359
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-172-063-21

Query Match 98.5%; Score 1256; DB 3; Length 323;
Best Local Similarity 98.7%; Pred. No. 1.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 144
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKEGDTL 120
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKEGDTL 204
QY 121 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 10
US-09-602-641-21
; Sequence 21, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
```

```
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-602-641-21

Query Match 98.5%; Score 1256; DB 4; Length 323;
Best Local Similarity 98.7%; Pred. No. 1.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 144
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKEGDTL 120
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKEGDTL 204
QY 121 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 11
US-09-085-305-6
; Sequence 6, Application US/09085305
; Patent No. 6191269
; GENERAL INFORMATION:
; APPLICANT: Pollock, Allan
; APPLICANT: Lovett, David H.
; TITLE OF INVENTION: Selective Induction of Apoptosis in
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,305
; FILING DATE: 29-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
```


ATTORNEY/AGENT INFORMATION:
NAME: Francis, Carol L
REGISTRATION NUMBER: 36,513
REFERENCE/DOCKET NUMBER: 6510/102US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-085-305-6

Query Match 98.5%; Score 1256; DB 3; Length 364;
Best Local Similarity 98.7%; Pred. No. 2.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPVWPT 60
DB 126 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPVWPT 185

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 186 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 245

QY 121 VNRLEKGDIFKEDGNILGKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 246 VNRLEKGDIFKEDGNILGKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 305

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 306 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 364

RESULT 12
US-09-417-197-129
Sequence 129, Application US/09417197
Patent No. 6518021
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE REFERENCE: 3759-0110P
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 129
LENGTH: 379
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: actin-binding-domain-EGFP fusion
US-09-417-197-129

Query Match 98.5%; Score 1256; DB 4; Length 379;
Best Local Similarity 98.7%; Pred. No. 2.2e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPVWPT 60
DB 141 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPVWPT 200

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 201 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 260

QY 121 VNRLEKGDIFKEDGNILGKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 261 VNRLEKGDIFKEDGNILGKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 320

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 321 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 379

RESULT 13
US-09-800-170-48
Sequence 48, Application US/09800170
Patent No. 6481667
GENERAL INFORMATION:
APPLICANT: Kinsella, Todd
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
FILE REFERENCE: A-68614-1/DJB/RMS/RMK
CURRENT APPLICATION NUMBER: US/09/800,170
CURRENT FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: US 60/187,130
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 90
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48
LENGTH: 434
TYPE: PRT
ORGANISM: Synechocystis PCC6803
US-09-800-170-48

Query Match 98.5%; Score 1256; DB 4; Length 434;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPVWPT 60
DB 196 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPVWPT 255

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 256 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 315

QY 121 VNRLEKGDIFKEDGNILGKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 316 VNRLEKGDIFKEDGNILGKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 375

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 376 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 434

RESULT 14
US-09-417-197-127
Sequence 127, Application US/09417197
Patent No. 6518021
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE REFERENCE: 3759-0110P
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 127
LENGTH: 442
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EGFP-RhoA fusion
US-09-417-197-127

Query Match 98.5%; Score 1256; DB 4; Length 442;
Best Local Similarity 98.7%; Pred. No. 2.8e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPVWPT 60

Db	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVWPPT	60
Qy	61	LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKFECDTL	120
Db	61	LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKFECDTL	120
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLGFVTAAGITLGMDELYK	239

RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPP-HSP27
US-09-513-783A-170

Query Match	98.5%	Score 1256;	DB 4;	Length 459;
Best Local Similarity	98.7%	Pred. No. 2.9e-127;	Indels 0;	Gaps 0;
Matches 236;	Conservative 1;	Mismatches 2;		
Qy	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVWPPT	60	
Db	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVWPPT	60	
Qy	61	LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKFECDTL	120	
Db	61	LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKFECDTL	120	
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180	
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180	
Qy	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLGFVTAAGITLGMDELYK	239	
Db	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLGFVTAAGITLGMDELYK	239	

Search completed: June 21, 2004, 16:04:00
Job time : 13.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds
(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-64G
Perfect score: 1275
Sequence: 1 MVSKEELFTGVVPILVELD.....VLLGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1267	99.4	239	9 US-09-887-784-4	Sequence 4, Appli
2	1267	99.4	239	12 US-10-296-953-4	Sequence 4, Appli
3	1267	99.4	363	14 US-10-270-223-6	Sequence 6, Appli
4	1267	99.4	893	14 US-10-257-909A-30	Sequence 30, Appli
5	1267	99.4	1132	14 US-10-257-909A-32	Sequence 32, Appli
6	1259	98.7	239	9 US-09-887-784-2	Sequence 2, Appli
7	1259	98.7	239	12 US-10-296-953-2	Sequence 2, Appli
8	1256	98.5	239	9 US-09-920-922-2	Sequence 2, Appli
9	1256	98.5	239	9 US-09-999-745-4	Sequence 4, Appli
10	1256	98.5	239	10 US-09-866-538-4	Sequence 4, Appli
11	1256	98.5	239	10 US-09-797-496B-2	Sequence 2, Appli
12	1256	98.5	239	10 US-09-794-308-4	Sequence 4, Appli
13	1256	98.5	239	10 US-09-865-291-4	Sequence 4, Appli
14	1256	98.5	239	12 US-10-457-982-3	Sequence 3, Appli
15	1256	98.5	239	14 US-10-121-258-13	Sequence 13, Appli

16	1256	98.5	239	14	US-10-221-461-7	Sequence 7, Appli
17	1256	98.5	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1256	98.5	239	14	US-10-177-390-2	Sequence 2, Appli
19	1256	98.5	239	14	US-10-338-411-3	Sequence 3, Appli
20	1256	98.5	239	15	US-10-370-570-4	Sequence 4, Appli
21	1256	98.5	239	15	US-10-389-640-3	Sequence 3, Appli
22	1256	98.5	259	14	US-10-314-861-11	Sequence 11, Appli
23	1256	98.5	281	12	US-09-931-232-1	Sequence 1, Appli
24	1256	98.5	288	14	US-10-314-861-37	Sequence 37, Appli
25	1256	98.5	293	14	US-10-314-861-35	Sequence 35, Appli
26	1256	98.5	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1256	98.5	295	14	US-10-314-861-39	Sequence 39, Appli
28	1256	98.5	299	14	US-10-314-861-33	Sequence 33, Appli
29	1256	98.5	305	14	US-10-314-861-31	Sequence 31, Appli
30	1256	98.5	308	14	US-10-033-717-35	Sequence 35, Appli
31	1256	98.5	311	14	US-10-314-861-29	Sequence 29, Appli
32	1256	98.5	320	14	US-10-338-411-11	Sequence 11, Appli
33	1256	98.5	320	15	US-10-389-640-11	Sequence 11, Appli
34	1256	98.5	323	12	US-10-457-982-21	Sequence 21, Appli
35	1256	98.5	323	14	US-10-338-411-7	Sequence 7, Appli
36	1256	98.5	323	14	US-10-338-411-13	Sequence 13, Appli
37	1256	98.5	323	15	US-10-389-640-7	Sequence 7, Appli
38	1256	98.5	323	15	US-10-389-640-13	Sequence 13, Appli
39	1256	98.5	324	14	US-10-314-861-16	Sequence 16, Appli
40	1256	98.5	345	14	US-10-338-411-5	Sequence 5, Appli
41	1256	98.5	345	15	US-10-389-640-5	Sequence 5, Appli
42	1256	98.5	346	14	US-10-338-411-9	Sequence 9, Appli
43	1256	98.5	346	15	US-10-389-640-9	Sequence 9, Appli
44	1256	98.5	359	14	US-10-033-717-33	Sequence 33, Appli
45	1256	98.5	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BIORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequoria Victoria
US-09-887-784-4

Query Match 99.4%; Score 1267; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.1e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLT	KFICTTGKLPVWPPT	60
Db	1	MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLT	KFICTTGKLPVWPPT	60
Qy	61	LVTGLSYGVQCFSRYPDHMKQDFFKSA	MPGVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Db	61	LVTGLSYGVQCFSRYPDHMKQDFFKSA	MPGVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Qy	121	VNRIELKIDFDKEDGNILGHKLEYNHN	YVIMADKQKNGIKVNFKIRHNIEDGSVQLA	180
Db	121	VNRIELKIDFDKEDGNILGHKLEYNHN	YVIMADKQKNGIKVNFKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPIGDGPVLLPDNHHYLS	QTSALSQDPNEKRDNHVLGFTVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDGPVLLPDNHHYLS	QTSALSQDPNEKRDNHVLGFTVTAAGITLGMDELYK	239

RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4
Query Match 99.4%; Score 1267; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.1e-123; Indels 0; Gaps 0;
Matches 238; Conservative 0; Mismatches 1;
QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
Db 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Query Match 99.4%; Score 1267; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.1e-123; Indels 0; Gaps 0;
Matches 238; Conservative 0; Mismatches 1;
QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
Db 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
RESULT 3
US-10-270-423-6
; Sequence 6, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: BioImage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPOUNDS AND TO EXTRACT QUANTITATIVE INFORMATION FROM THE INTERACTIONS BY FLUORESCENCE REDISTRIBUTION.
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6
Query Match 99.4%; Score 1267; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 1.9e-123; Indels 0; Gaps 0;
Matches 238; Conservative 0; Mismatches 1;
QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
Db 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30
Query Match 99.4%; Score 1267; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 7e-123; Indels 0; Gaps 0;
Matches 238; Conservative 0; Mismatches 1;
QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
Db 655 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 714
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 715 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 775 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 835 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 893
RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0

QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
Db 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30
Query Match 99.4%; Score 1267; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 7e-123; Indels 0; Gaps 0;
Matches 238; Conservative 0; Mismatches 1;
QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 60
Db 655 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVWPPT 714
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 715 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 775 VNRLEKIGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 835 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 893
RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0

```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.4%; Score 1267; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 9.8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
Db 894 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 953

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 954 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 1014 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGFTVTAAGITLGMDELYK 239
Db 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGFTVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          98.7%; Score 1259; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 7.2e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGFTVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGFTVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          98.5%; Score 1256; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
```

QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 9

US-09-999-745-4
; Sequence 4, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-999-745-4

Query Match 98.5%; Score 1256; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTLPKICTTGKLPVPWPT 60
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY :21 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB :21 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY :181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB :181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 10

US-09-866-538-4
; Sequence 4, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 4
; LENGTH: 239
; TYPE: PRT

; ORGANISM: Aequorea victoria
US-09-866-538-4
Query Match 98.5%; Score 1256; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTLPKICTTGKLPVPWPT 60
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
RESULT 11
US-09-797-496B-2
; Sequence 2, Application US/09797496B
; Publication No. US20030049597A1
; GENERAL INFORMATION:
; APPLICANT: Simon, Sanford M.
; APPLICANT: Chen, Yu
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof
; FILE REFERENCE: 600-1-267
; CURRENT APPLICATION NUMBER: US/09/797,496B
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described
; OTHER INFORMATION: in specification
US-09-797-496B-2

Query Match 98.5%; Score 1256; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTLPKICTTGKLPVPWPT 60
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 12

US-09-794-308-4
; Sequence 4, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA

APPLICANT: TSIEN, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match
Best Local Similarity 98.5%; Score 1256; DB 10; Length 239;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
Sequence 4, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEN, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REGEN1550
CURRENT APPLICATION NUMBER: US/09/865,291
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match
Best Local Similarity 98.5%; Score 1256; DB 10; Length 239;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US2003021265A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Acsushi
APPLICANT: Lloplis, Juan
APPLICANT: Wachter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR APPLICATION NUMBER: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match
Best Local Similarity 98.5%; Score 1256; DB 12; Length 239;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 15
US-10-121-258-13
Sequence 13, Application US/10121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP1
CURRENT APPLICATION NUMBER: US/10/121,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24

Sun Jun 27 18:27:36 2004

; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match 98.5%; Score 1256; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTTLKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTTLKFICTTGKLPVPWPT 60
Qy 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSXDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSXDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:25
Job time : 35.7778 secs

A;Contents: annotation: X-ray crystallography, 1.9 angstroms
C;Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
A;Genetics:
A;Gene: GFP
A;Introns: 69/3; 167/3
A;Superfamily: green-fluorescent protein
C;Keywords: chromoprotein; luminescence
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 96.8%; Score 1234; DB 1; Length 238;
Best Local Similarity 96.6%; Pred. No. 6.4e-96;
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 VSKGEELFTGVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

Qy 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFPKDGNKYKTRAEVKFEGDTLV 121
Db 61 VTFISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFPKDGNKYKTRAEVKFEGDTLV 120

Qy 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKYKFNIRHNIEDGSVQLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKYKFNIRHNIEDGSVQLAD 180

Qy 182 HYQONTPIGDPVLLPNHYLSTQSALSADPNKRDHMLLVFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPNHYLSTQSALSADPNKRDHMLLVFVTAAGITLGMDELYK 238

RESULT 2
H72228
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;References: A72200; MUID:99287316; PMID:10360571
A;Accession: H72228
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-785 <ARN>
A;Cross-references: GB:AE001806; GB:AE000512; NID:94982196; PIDN:AAD36691.1; PID:g498219
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1624

Query Match 8.3%; Score 106; DB 2; Length 785;
Best Local Similarity 20.2%; Pred. No. 0.75;
Matches 47; Conservative 33; Mismatches 69; Indels 84; Gaps 8;

Qy 3 SKGEELFTGVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTLV 62
Db 15 NEGFSFEGTVPGVVQAD-----LVRKGLLPHPY----- 43

Qy 63 TGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFPKDGNKYKTRAEVKFEGDTLVN 122
Db 44 VGMN-----EDLFKIEDIEWIYEFPEFKEDVKEGERVDLVFEGVDTL 88

Qy 123 RIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKYKFNIRHNIEDGSVQLADH 182
Db 89 DVLNGVYL--GSTEDMFIYRFDTNVL-----KEKNHLYK-----SPIRVPKT 134

Qy 183 YQONTPIGDPVLLPNHYLSTQSALSADPNKRDHMLLVFVTAAGITLGM 235
Db 135 LEQNYGVLGSP-----EDF-----IRGYIRKQSYGYHD 163

RESULT 3
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C;Species: Haemophilus influenzae
A;Variety: type b
C;Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C;Accession: JC4078
R;Flack, F.S.; Loomore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A;Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl
A;Reference number: JC4078; MUID:95255676; PMID:7737523
A;Accession: JC4078
A;Molecule type: DNA
A;Residues: 1-797 <FLA>
A;Cross-references: GB:U13961; NID:9537447; PIDN:AAA85645.1; PID:g537448
A;Experimental source: type b
C;Superfamily: protective surface antigen D-15
C;Keywords: surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 7.3%; Score 92.5; DB 2; Length 797;
Best Local Similarity 22.3%; Pred. No. 10;
Matches 49; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

Qy 64 GLSYGVQCFSRYPDHMKQHD-----FKSAMPEGYVQ-----RT 98
Db 426 GIGYGTSGISYQASVKQDNFLGTGAAYSIAGTKNDYGTSVNLGYTEPYFTKDGVS LGN 485

Qy 99 IFFKDDGNKYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYN 147
Db 486 VFFENYDNSKSDTSSNYKRTYGSNVTL-GFPVNNNSYVGLGHTYKISNFALEYN-- 542

Qy 148 SHNVIYIMADKQK-NGIKYKFNIRHNIEDGSVQLADHYQQ-----NTPIGDGPVL 195
Db 543 -RNLYIQSMKFKNGIKTN-----DPDFSGWYNSLNRGYPTKGVKASLG-GRVT 592

Qy 196 LP--DNHYLSTQSALSADPNKRDHMLLVFVTAAGITLG 233
Db 593 IFGSDNKKYKLSADVQGFYPLDRDHLWVVSASAKAGYANG 632

RESULT 4
F64102
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C;Accession: F64102
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
; D.M.; Brandon, R.C.; Pine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: F64102
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-808 <TIGR>
A;Cross-references: GB:L42023; TIGR:HI0917
C;Superfamily: protective surface antigen D-15
C;Keywords: surface antigen

Query Match 7.3%; Score 92.5; DB 2; Length 808;
Best Local Similarity 22.3%; Pred. No. 10;
Matches 49; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

Qy 64 GLSYGVQCFSRYPDHMKQHD-----FKSAMPEGYVQ-----RT 98
Db 439 GIGYGTSGISYQASVKQDNFLGTGAAYSIAGTKNDYGTSVNLGYTEPYFTKDGVS LGN 498

Qy 99 IFFKDDGNKYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYN 147

RESULT 6
G95153
neuramin

Sun Jun 27 18:27:37 2004

Db 123 SVVNALEWLVIEIKRDGMYEQRPENGGKPTSTILEKKGKTRQGTGTTIHFKDPDPTVFSTT 182
QY 106 --NYKTRAEVKFGDTLVNRIELKGDIDFKEDGNILGHKLEYNVSHNVYIMADK----- 157
Db 183 NFNVETLSERLEAAFLKGLKTELVDLRDDTKVEFH-YEDGIKAPVEYINEDKETLHPV 241
QY 158 -----QKNGIKVNFKIRHNIEDQSVQADHYQONTPIGDGPVLLPDNHYLSTQSALSADP 212
Db 242 VFPNGSGNIEIEFAQFN--DGYTENVLSPVNVVTKDG-----GTHELGAKTAMTRAV 294
QY 213 NE 214
Db 295 NE 296
RESULT 8
H64102
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)
N:Alternate names: leucyl-tRNA synthetase
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
C:Accession: H64102
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 265, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter,
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-861 <TIGR>
A:Cross-references: GB:U32774; GB:L42023; NID:gi573942; PIDN:AAC22581.1; PID:gi573943; T
C:Genetics:
A:Gene: leuS
C:Superfamily: leucine-tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis
Query Match 6.9%; Score 88.5; DB 2; Length 861;
Best Local Similarity 24.1%; Pred. No. 25;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVWPVTLVGLSYGVCFSRYPDHMKQ-HDFEKSAMPEGYVOERTIFFKD----- 103
Db 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQDF-----EPAQKYSLPKQVIAPLA 364
QY 104 DGNKTRAEVKFGDTLVNRIELKGDIDFKEDGNILGHKLEYNVSHNVYIMADK-QKNGI 162
Db 365 DEEDLTQKQFVEHGKLVNSDEFQKNF--DQAFNG-----IADKLEKLV 408
QY 163 ---KVNFKIRH-----NIEDGSVOLADHYQONTPIGDGPVLLPDNHYL- 202
Db 409 GKRQVNYLRDQVSRQRYWYGAPIMLTLENGDVVPA-----PMEDLPILPEDVYMD 461
QY 203 STQSALSADPN 213
Db 462 GVKSPINADPN 472
RESULT 9
E70390
iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus
N:Contains: L-cysteine sulfurtransferase (EC 2.8.1.1-)
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999
C:Accession: E70390
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70390
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-370 <AQF>
A:Cross-references: GB:AB000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE000657;
A:Experimental source: strain VF5
C:Genetics:
A:Gene: nifs1
C:Superfamily: nitrogen fixation protein nifs
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase
F:195/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted
F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted
Query Match 6.9%; Score 87.5; DB 2; Length 370;
Best Local Similarity 25.4%; Pred. No. 9.9;
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;
QY 4 KGEELFTGW----PILVELD---GDVNGHKP-SVSGEG---EGDATYQKLTILKICT 50
Db 164 KGVPLTTDAVQAIGKIPTELKNISVATPSGHKFAIKSGFLYISDEANYEPLIVGGQE 223
QY 51 TQKLP-----VPWPTLVLTGLSYGVCFSRYPDHMKQ-HDFEKSAMPEGYVOERTIFFKD 104
Db 224 NGRSGTENVVGILSLAKALBIIIVSNFSRYQQLKLRDLFENLLEA-LPDAQIVGKA 282
QY 105 GNYKTRAEV---KPGEDTLVNRIELKGDIDFKEDGNILGHKLEYNVSHNVYIMADKQNG 161
Db 283 ERSPSISSVIMPKFPGAIVNKLSEKGIYCTSGSACLSGEYEPNKHMLKMGFSQEKALRM 342
QY 162 IKVNFKIRHNIED 174
Db 343 VRFSFGLLNKEE 355
RESULT 10
AI2140
polyketide synthase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AI2140
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takaazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anati
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2518 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074379.1; PID:gi17131773; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: air2680
Query Match 6.9%; Score 87.5; DB 2; Length 2518;
Best Local Similarity 24.5%; Pred. No. 1.2e+02;
Matches 48; Conservative 24; Mismatches 69; Indels 55; Gaps 11;
QY 48 ICTTGKLPVWPVTLVGLSYGVCFSRY---PDH-----MKQHDFFK----- 86
Db 1537 IVLEGTIPSIIWDLIFGLTEGWRWFQDQLRPHPLISTTAWHSLKLTHTDFTNIVNITPD 1596
QY 87 SAMPEGYVOERTIFFKD-----DGNKTRAE-----VKFEGDTLVNRIELKIDFKEDG 135
Db 1597 SILPEALAQOSVIAVQNTPPQSPYKREGSECIITDLPANGAIIHQQLKLPSTFKPSD 1656
QY 136 NILGHKLEYNVNS--HNVYIMADKQNGIKVNFKIRHNIEDQSVQADHYQONTPIGDGP 193
Db 1657 NI-----AEFPHQSIKHIYI-ACQDNITECN-NILHLVQ--TLIKTQHY----- 1698
QY 194 VLLPDNHYLSTQSALS 209

Db 1699 ---PINLWLVTQGAIS 1711

RESULT 11

D71614

hypothetical protein PFB0460c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: D71614

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1128-1132, 1998

A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

C:Accession: D71614

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2573 <GAR>

A:Cross-references: GB:AE001362; GB:AE001362; NID:g3845188; PIDN:AAC71881.1; PID:g384519

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;

Best Local Similarity 26.2%; Pred. No. 1.2e+02;

Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VQERTIFFKD--DGNVKTAEVKEFGDTLVNRTELKGDIDPKEDGNILGHKLEYN--YN5H 149

Db 126 LKKEITLKKDKSGSDNDWDEISLFKDDMWDDKELK--DFEKSLLKIKKKEVNFYINQM 183

QY 150 NVYIMADKQKNGIKVNFKIRHNIEDSGVQLADHYQQNTPIGDGPVLLPNHLYLSTQSALS 209

Db 184 NLHIKENKKDEKKNKHNNDDNNM---IYYKNI---DKTHYLDNNVVVHILNDIN 236

QY 210 KDPNEKRDHM 219

Db 237 TVLKRERYDM 246

RESULT 12

KXRTCI

proteine convertase 1 (EC 3.4.21.93) precursor - rat

N:Alternate names: furin homolog PC1; kexin homolog PC1; prohormone cleavage enzyme; pro

C:Species: Rattus norvegicus (Norway rat)

C>Date: 31-Dec-1992 #8sequence_revision 31-Dec-1992 #text_change 19-May-2000

C:Accession: A41556; S27361; S36358

R:Bloomquist, B.T.; Eipper, B.A.; Mains, R.E. Mol. Endocrinol. 5, 2014-2024, 1991

A>Title: Prohormone-converting enzymes: regulation and evaluation of function using anti

A:Reference number: A41556; MUID:92168040; PMID:1791845

A:Accession: A41556

A:Molecule type: mRNA

A:Residues: 1-752 <BLO>

A:Cross-references: GB:M76705; NID:G203508; PIDN:AAA40945.1; PID:G203509

R:Hakes, D.J.; Birch, N.P.; Mezey, A.; Dixon, J.E. Endocrinology 129, 3053-3063, 1991

A>Title: Isolation of two complementary deoxyribonucleic acid clones from a rat insulin

to endocrine and neuroendocrine tissues in rats.

A:Reference number: S27361; MUID:92063860; PMID:1954888

A:Accession: S27361

A:Molecule type: mRNA

A:Residues: 1-513,'A',515-752 <HAK>

A:Cross-references: EMBL:M83745

R:Hakes, D.J.; Birch, N.P.; Mezey, A.; Dixon, J.E. submitted to the EMBL Data Library, February 1992

A:Reference number: S36358

A:Accession: S36358

A:Molecule type: mRNA

A:Residues: 1-366,'T',367-513,'A',515-752 <HAK2>

A:Cross-references: EMBL:M83745; NID:g205062; PIDN:AAA41476.1; PID:g205063

C:Comment: This protein is a member of a family of subtilisin-like proteinases responsible

C:Comment: This protein lacks a classical hydrophobic transmembrane segment but may asss

C:Superfamily: kexin; subtilisin homology

C:Keywords: glycoprotein; hydrolase; serine proteinase

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-110/Domain: propeptide #status predicted <PRO>

F:111-752/Product: prohormone-processing proteinase PC1 #status predicted <MAT>

F:158-396/Domain: subtilisin homology <SBT>

F:158-396/Domain: subtilisin homology <SBT>

F:173,401.645/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.8%; Score 87; DB 1; Length 752;

Best Local Similarity 25.6%; Pred. No. 28;

Matches 53; Conservative 20; Mismatches 70; Indels 64; Gaps 10;

QY 18 ELDGDVNGHK----FVSVGEGBDQATYGLTKLTKFICTTGKLP-----VPMPTLVLTGLSYG 68

Db 540 ERDTSFNGFKNDWFMVHTWGENPV--GTWTLKVTDMSGRMQNEGRVNVWKLILHGT--- 594

QY 69 VOCFSRYPDHMKQHDFFKSKAMPEGYVQERTIFFKQDGNKYKTRAEVKEFGDTLVNRTELKG 128

Db 595 ----SSQPEHMQ-----PRVYT-----SYNTVQNDRRGVKMKVNVVEEKP 631

QY 129 IDPKEDGNILGHKLEYNVNSHNVYIMADKQKNGIKVNFKIRHNIEDSGVQLADHYQQNTP 188

Db 632 TQNSLNGNLVFK---NSSSSVEDRDRDEQVGAPSKAMLR-----LLQSAFSKNTP 680

QY 189 IGDGPVLLPDNHLYLSTQSALSADPNK 215

Db 681 -----SKQS--SKIPSAK 691

RESULT 13

E82590

leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C:Accession: E82590

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: E82590

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-887 <SIM>

A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN0013

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; AJ Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Perro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme J.D.; Junqueira, M.B.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaka M.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zaf

Query Match 6.8%; Score 86.5; DB 2; Length 887;

Best Local Similarity 22.2%; Pred. No. 38;

Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPMPTLVLTGLSYGVQCFSRYPDHMKQHDFFKSKAMPEGYVQERTIFFKQDGNY-- 107

Db 329 TNEQLPV-VWANFVLMAYGTGAVMVFQDQDQDEF--ANKYGLPIRQVIALKEPKNQDE 385

```

F;5-77/Domain: lipoyl/biotin-binding homology <LPB>
F;117-145/Region: beta-alpha-beta FAD nucleotide-binding fold
F;119-561/Domain: dihydrolipoamide dehydrogenase homology <LBD>
F;287-315/Region: beta-alpha-beta NAD nucleotide-binding fold
F;153-158/Disulfide bonds: redox-active #status predicted

Query Match      6.7%; Score 86; DB 1; Length 578;
Best Local Similarity 23.5%; Pred. No. 24;
Matches 54; Conservative 39; Mismatches 85; Indels 52; Gaps 12;

QY      10  TGVVPTLVLDGVDNGHKFSVSGEGSDATYGLTLK-----FICTGKLPVWPPTLVT 63
        255  TGSMPPIPEI-----GNKLS-----GVIDST-GALSLESNPESIALIGGGVIGVEFASIFN 305

DB      64  GLSYGVQCFSRYPDHMQHDFFKSAMPEGVQERTIFFKDDGNKYKPAEYKFGEDTLVNR 123
        306  SLGCKVSIIEMLPHILPMDREISEI-----AKAKLIIRDGININNN 346

QY      124  IELKGIDFKEDG---NILGHLEVYVNSHNVYIMADKQK--GIKNFKIRHNIEDGSVQ 178
        347  CKVTRIEQGEDGLKVFIGDKGEESIDVEKVLAVGRSNIEGLDVE-KIGVKTEGGSII 405

QY      179  LADHYQONT----PIGD--GPVLLPNHNYLSTQALSKD-----PNEKRDH 218
        406  VNDKMEINVEGIYAGDCTCKIMLA--HVASDQGVAAENIMGQNKXNDY 453

Search completed: June 21, 2004, 16:01:54
Job time : 12.3333 secs

```

RESULT 15

I40794
dihydrolipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum
N;Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase
hydrogenase complex chain E3; S-complex 50K chain
C;Species: Clostridium magnum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: I40794
R;Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuchel, A.
J. Bacteriol. 176, 3614-3630, 1994
A;Title: Biochemical and molecular characterization of the Clostridium magnum acetoind
A;Reference number: I40789; MUID:94266715; PMID:8206840
A;Accession: I40794
A;Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A;Residues: 1-578 <KRU>
A;Cross-references: GB:L31844; NID:g4723324; PIDN:AAA21748.1; PID:g472330
C;Function:
A;Description: catalyzes the oxidation of dihydrolipoamide to lipoamide using NAD
A;Pathway: acetoind dehydrogenase enzyme system
C;Superfamily: Alkaligenes dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenas
C;Residue: nnn..oxidoreductase..redox-active disulfide

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-64G

Perfect score: 1275

Sequence: 1 MYSKGSELTGVVPLVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1240	97.3	238	1	GFP_AEQVI
2	94.5	7.4	1603	1	VIT5_CABEL
3	92.5	7.3	795	1	D152_HAEIN
4	92.5	7.3	797	1	D151_HAEIN
5	92.5	7.3	941	1	GN_BACS6
6	91.5	7.2	533	1	CUP51_CANGA
7	91.5	7.2	793	1	D153_HAEIN
8	88.5	6.9	861	1	SYL_HAEIN
9	87	6.8	752	1	NECI_RAT
10	86.5	6.8	879	1	SYL_XYLFA
11	86	6.7	357	1	TRM_CAMJE
12	85.5	6.7	217	1	EXPI_ERWCA
13	85.5	6.7	886	1	ITH3_MESAU
14	85	6.7	439	1	SYG2_DISOM
15	84.5	6.6	468	1	GLNA_AZOCA
16	84.5	6.6	501	1	AMPA_WIGBR
17	84.5	6.6	613	1	PEPF_WICPU
18	84.5	6.6	788	1	DPO1_HPBHE
19	84.5	6.6	860	1	SYL_SALT1
20	84.5	6.6	860	1	SYL_SALT2
21	84	6.6	504	1	YC03_KLEPN
22	83.5	6.5	879	1	SYL_XYLFT
23	83.5	6.5	1164	1	BAG_STRAG
24	83	6.5	782	1	YM62_YEAST
25	82.5	6.5	369	1	FORL_BUCAP
26	82.5	6.5	658	1	ADAS_HUMAN
27	82.5	6.5	1224	1	COPA_HUMAN
28	82	6.4	353	1	HIS7_BUCAI
29	82	6.4	461	1	PSBC_CVAPA
30	82	6.4	774	1	AMY2_SCHPO
31	81.5	6.4	400	1	FMDC_METTM
32	81.5	6.4	400	1	FMDC_METTM
33	81.5	6.4	589	1	SYD_HAEDU

RESULT 1
GFP_AEQVI STANDARD; PRT; 238 AA.
ID P42212; Q17104; Q27903;
AC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92175527; PubMed=1347277;
RA Praher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
Cormier M.J.;
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein.";
RL Gene 111:229-233(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94185810; PubMed=8137953;
RA Inouye S., Tsuji F.I.;
RT "Aequorea green fluorescent protein. Expression of the gene and
RT fluorescence characteristics of the recombinant protein.";
RL FEBS Lett. 341:277-280(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
RT protein by modification of its codon usage.";
RL Plant Mol. Biol. 33:989-999(1997).
RN [4]
RP CHROMOPHORE.
RX MEDLINE=93192221; PubMed=8448132;
RA Cody C.W., Praher D.C., Westler W.M., Prendergast F.G., Ward W.W.;
RT "Chemical structure of the hexapeptide chromophore of the Aequorea
RT green-fluorescent protein.";
RL Biochemistry 32:1212-1218(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=96355665; PubMed=8703075;
RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,
Remington S.J.;
RT "Crystal structure of the Aequorea victoria green fluorescent
RT protein.";
RL Science 273:1392-1395(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=98294543; PubMed=9631087;
RA Yang F., Moss L.G., Phillips G.N. Jr.;
RT "The molecular structure of green fluorescent protein.";
RL Nat. Biotechnol. 14:1246-1251(1996).

ALIGNMENTS

34 81.5 6.4 859 1 SYL_SHEON
35 81.5 6.4 1603 1 VIT5_CABEL
36 81 6.4 336 1 YD48_METJA
37 81 6.4 682 1 PRC_ECOLI
38 81 6.4 753 1 NECI_MOUSE
39 80.5 6.3 393 1 TRMB_HELPY
40 80.5 6.3 538 1 GRBE_RAT
41 80.5 6.3 953 1 YNM7_YEAST
42 80.5 6.3 2222 1 DPOE_YEAST
43 80 6.3 689 1 AC2L_HUMAN
44 80 6.3 737 1 OPT1_DROME
45 80 6.3 874 1 SLAP_BACLI

Q8ehp4 shewanella
P06125 caenorhabdi
Q58743 methanococc
P23865 escherichia
P21662 mus musculus
O25443 helicobacte
O88900 rattus norv
P53917 saccharomyc
P21951 saccharomyc
Q9nub1 homo sapien
P91679 drosophila
P49052 bacillus li

[7]
X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
MEDLINE=98455509; PubMed=9782051;
Wachter R.M., Elsiger M.A., Kallio K., Hanson G.T., Remington S.J.,
"Structural basis of spectral shifts in the yellow-emission variants
of green fluorescent protein.";
Structure 6:1267-1277(1998).
[8]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
MEDLINE=99238303; PubMed=10220315;
Elsiger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.,
"Structural and spectral response of green fluorescent protein
variants to changes in pH.";
Biochemistry 38:5296-5301(1999).
-1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the
blue chemiluminescence of the protein aequorin into green
fluorescent light by energy transfer. Fluoresces in vivo upon
receiving energy from the Ca(2+)-activated photoprotein aequorin.
Absorbs light maximally at 395 nm and exhibits a smaller
absorbance peak at 470 nm. The fluorescence emission spectrum
peaks at 509 nm with a shoulder at 540 nm.
-1- SUBUNIT: Monomer.
-1- TISSUE SPECIFICITY: Photocytes.
-1- PFM: Contains a covalently attached chromophore, which is composed
of modified amino acid residues. The chromophore is formed upon
cyclization of the residues Ser-dehydroTyr-Gly.
-1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
chimeric proteins of GFP linked to other proteins where it
functions as a fluorescent protein tag. GFP tolerates N- and C-
terminal fusion to a broad variety of proteins. It has been
expressed in bacteria, yeast, slime mold, plants, Drosophila,
zebrafish, and in mammalian cells. As a noninvasive fluorescent
marker in living cells, it allows for a wide range of applications
where it may function as a cell lineage tracer, reporter of gene
expression, or as a measure of protein-protein interactions.
-1- DATABASE: NAME-Protein Spotlight;
NOTE=Issue 11 of June 2001;
WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M62654; AAA27722.1; -;
EMBL; M62653; AAA27721.1; -;
EMBL; L29345; AAA58246.1; -;
EMBL; X96418; CAA65278.1; -;
PIR; JS0692; JQ1514.
PDB; 1B9C; 17-NOV-00.
PDB; 1BFP; 07-JUL-97.
PDB; 1CAF; 14-JUN-00.
PDB; 1EMA; 08-NOV-96.
PDB; 1EMB; 16-JUN-97.
PDB; 1EMC; 20-AUG-97.
PDB; 1EME; 20-AUG-97.
PDB; 1EMF; 20-AUG-97.
PDB; 1EMG; 12-MAY-99.
PDB; 1EMK; 20-AUG-97.
PDB; 1EML; 20-AUG-97.
PDB; 1EMM; 20-AUG-97.
PDB; 1F09; 17-NOV-00.
PDB; 1FOB; 17-NOV-00.
PDB; 1GFL; 11-JAN-97.
PDB; 1HCJ; 15-JAN-02.
PDB; 1HUY; 04-JUL-01.
PDB; 1JBY; 07-JAN-03.
PDB; 1JBZ; 07-JAN-03.
PDB; 1KPS; 28-AUG-02.
PDB; 1KYP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.
DR PDB; 1KYS; 10-APR-02.
DR PDB; 1YFP; 28-OCT-98.
DR PDB; 2EMD; 20-AUG-97.
DR PDB; 2EMN; 20-AUG-97.
DR PDB; 2EMO; 20-AUG-97.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR00786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
KW Luminescence; 3D-structure.
FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).
FT MOD_RES 66 66 2,3-DIHYDROXYROSINE.
FT VARIANT 100 100 F -> Y.
FT VARIANT 108 108 T -> S.
FT VARIANT 141 141 L -> M.
FT VARIANT 219 219 V -> I.
FT CONFLICT 2 2 S -> G (IN REF. 3).
FT CONFLICT 25 25 H -> Q (IN REF. 2).
FT CONFLICT 80 80 Q -> R (IN REF. 3).
FT CONFLICT 157 157 Q -> P (IN REF. 2).
FT CONFLICT 172 172 E -> K (IN REF. 2).
FT HELIX 4 8
FT STRAND 12 22
FT TURN 23 24
FT STRAND 25 36
FT TURN 37 40
FT STRAND 41 48
FT TURN 49 50
FT TURN 57 60
FT TURN 61 63
FT HELIX 69 71
FT STRAND 73 73
FT HELIX 76 81
FT HELIX 83 86
FT TURN 87 90
FT STRAND 92 100
FT TURN 101 102
FT STRAND 105 115
FT TURN 116 117
FT STRAND 118 128
FT TURN 132 133
FT TURN 135 139
FT STRAND 141 141
FT STRAND 148 155
FT TURN 156 159
FT STRAND 160 171
FT TURN 172 173
FT STRAND 176 187
FT STRAND 199 208
FT TURN 211 212
FT STRAND 217 227
SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6E05 CRC64;
Query Match 97.3%; Score 1240; DB 1; Length 238;
Best Local Similarity 97.9%; Pred. No. 1.3e-95;
Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Qy 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDGYGKLTFLKFCITGKLPVPEPTL 61
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDGYGKLTFLKFCITGKLPVPEPTL 60
Qy 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIPFKDDGNKTRAEVKEFGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIPFKDDGNKTRAEVKEFGDTLV 120
Qy 122 NRTELKIDFKEDGNLTGLHKLKLVNYSNHYINADKKNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRTELKIDFKEDGNLTGLHKLKLVNYSNHYINADKKNGIKVNFKIRHNIEDGSVOLAD 180
Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMVLLGFVTAAGITLGMDELYK 239
Db 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMVLLGFVTAAGITLGMDELYK 239

Db 426 GIGYTESGISYQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTTKDGVSLGNN 485
QY 99 IPFKDGNVYKTRAEVKEFGDGLVNRLELKGIDFKEDGNI---LGH-----KLEYNYN 147
Db 486 VFPENYDNSKSDTSNRYKRTTYGSNVTL-GFPVNNNSYYVGLGHTYNNKISNFALEYN-- 542
QY 148 SHNVYIMADKQK-NGIKVNFKIRHNIEDGSVQLADHYQQ-----NTPIGDGPVL 195
Db 543 -RNLYIQSMKFKNGIKTN-----DFDFSGWYNSLNRGYFTTKGVKASLG-GRVT 592
QY 196 LP--DNHLYSTQSALSADPNKPKRDHMLVLLGFTVTAAGITLG 233
Db 593 IPGSDNKKYKLSADVQGFYPLDRDLHLWVVSASAKASAGYANG 632

RESULT 4

D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RT Haemophilus influenzae.";
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Minna / Serotype B, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: Belongs to the surface antigen D15 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U13961; AAA85645.1; -;
CC EMBL; U60832; AAB61974.1; -;
CC EMBL; U60833; AAB61976.1; -;
CC PIR; JC4078; JCA078.
CC InterPro; IPR00184; Bac_surfAg_D15.
CC Pfam; PF01103; Bac_surface_Ag1.
CC Antigen; Outer membrane; Signal.
FT SIGNAL 1 19
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF18 CRC64;

Query Match 7.3%; Score 92.5; DB 1; Length 797;
Best Local Similarity 22.3%; Pred. No. 4.9;
Matches 49; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

64 GLSYGVQCFGRYPDHMKQHDF-----FKSAMPEGYVQE-----RT 98

Db 426 GIGYTESGISYQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTTKDGVSLGNN 485
QY 99 IPFKDGNVYKTRAEVKEFGDGLVNRLELKGIDFKEDGNI---LGH-----KLEYNYN 147
Db 486 VFPENYDNSKSDTSNRYKRTTYGSNVTL-GFPVNNNSYYVGLGHTYNNKISNFALEYN-- 542
QY 148 SHNVYIMADKQK-NGIKVNFKIRHNIEDGSVQLADHYQQ-----NTPIGDGPVL 195
Db 543 -RNLYIQSMKFKNGIKTN-----DFDFSGWYNSLNRGYFTTKGVKASLG-GRVT 592
QY 196 LP--DNHLYSTQSALSADPNKPKRDHMLVLLGFTVTAAGITLG 233
Db 593 IPGSDNKKYKLSADVQGFYPLDRDLHLWVVSASAKASAGYANG 632

RESULT 5

GUN_BACS6 STANDARD; PRT; 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Alkaline cellulase).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=2230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT "Molecular cloning and nucleotide sequence of a gene for alkaline
RT cellulase from Bacillus sp. KSM-635.";
RL J. Gen. Microbiol. 136:1327-1334(1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27420; AAA22304.1; -;
CC PIR; S29043; S29043.
CC PDB; 1G01; 31-DEC-02.
CC PDB; 1G0C; 31-DEC-02.
CC InterPro; IPR005086; CBM_17_28.
CC InterPro; IPR008979; Gal_bind_like.
CC InterPro; IPR001547; Glyco_hydro_5.
CC InterPro; IPR001119; SLH.
CC Pfam; PF03424; CBM_17_28; 2.
CC Pfam; PF00150; cellulase; 1.
CC Pfam; PF00395; SLH; 3.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PROSITE; PS01072; SLH_DOMAIN; 2.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat;
CC 3D-structure. 1 29
FT SIGNAL 30 941 ENDOGLUCANASE.
FT CHAIN 40 99 SLH 1.
FT DOMAIN 100 151 SLH 2.
FT DOMAIN 152 225 SLH 3.
FT ACT_SITE 373 373 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 485 485 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 941 AA; 104628 MW; BEA2AC3B169BFADA CRC64;

```
Query Match      7.3%; Score 92.5; DB 1; Length 941;
Best Local Similarity 20.2%; Pred. No. 6;
Matches 49; Conservative 34; Mismatches 60; Indels 99; Gaps 11;

QY 16 LVLDGVDVNGKFSVSGEGSDATYGLTLKFKICTTGKLPVWPVTLVTGLS-YGVQCFSR 74
DB 240 LVSENG-----QUTLAGE---DGT-----PVQLRGMSTHGLQWFG- 271

QY 75 YPDHMKQHDFFKSGAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTLVNRIELKIGDIFKED 134
DB 272 --EIVNENAFVLSNDWGSNIRLAWYIGENGATNPEVK--DLVYEGIELA----- 319

QY 135 GNILGHKLEYNHNVNHYIMADKQKGIKYNFKIRNIEDGSV-----QLADHYQOQNTPIG 190
DB 320 -----FEHDMYIVTDWH--VHAPGDPRADVSGAYDFEEIADHYKDH---- 360

QY 191 DGPVLDPDHYLSTQSALSXDPN-----EKRDHMYL 221
DB 361 -----PKNHYIILWELANSPNNGGPGLTNDEKGEAVKEVAPIVEMLRKGDNMIL 414

QY 222 LG 223
DB 415 VG 416

RESULT 6
CP51_CANGA
ID CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYP51) (P450-LIA1) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIA1) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Obtusifolliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
DR EMBL; L40389; AAB02329.1; -;
DR EMBL; S75389; AAB32679.1; -;
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
KW Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match      7.2%; Score 91.5; DB 1; Length 533;
Best Local Similarity 22.3%; Pred. No. 3.7;
Matches 46; Conservative 32; Mismatches 79; Indels 49; Gaps 10;

QY 25 GHKFSVS---GEGEGDATYCKLTCLKFKICTGKLPVWPVTLVTGLSYGVQCFSPYDHF--M 79
DB 109 GHEFIFNAKLADVSAAEAYSHLTT-----PVFGKGVY--DC-----PNHRLM 149

QY 80 KOHDFPKSAM-PEGYV-----OERTIFFKDDGNKYKTRAEVKEGDTLVNRIELKIGDIF 131
DB 150 EQKKFVKGALTKEAFVRYVPLTAEIYKYFRSNKFNINENSGIVDVVMSQSEM--TIF 207

QY 132 KEDGNILGHKLEYNHNVNHYIMADKQKGIKYNFKIRNIEDGSVQLADHYQOQNTPIGD 191
DB 208 TASRSLGLKEMRKLDTPFAYLSDLDKGFTPINF-VFPNLPLEHYKRDHQAQAIS--- 263

QY 192 GPVLLPDNHYLSTQSALSXDPNKRD 217
DB 264 -----GTYSMLIKERREKND 278

RESULT 7
D153_HAEIN
ID D153_HAEIN STANDARD; PRT; 793 AA.
AC O32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RX MEDLINE=97427952; PubMed=9284140;
RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
```

```

DR EMBL: J060834; AAB61977.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac surface Ag; 1.
DR Antigen: Outer membrane; Signal.
KW SIGNAL
FT CHAIN 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 518FD203680A14 CRC64;

Query Match
Best Local Similarity 7.2%; Score 91.5; DB 1; Length 793;
Matches 50; Conservative 28; Mismatches 79; Indels 63; Gaps 11;

QY 54 GUSYGVCFSRYEDHMKQDF-----FKSAMPEGVQVE-----RT 98
DQ 426 GIGYGTGESISQTSIKQNFGLTGAASVIACTKNDYGTISVNLGTEPYFTKDGVSIGGN 485
QY 99 IFFKDGNYKTRAEVKFEGDTLVNRIELKGDIFKEDGNI---LGH-----KLEYNYN 147
DQ 486 IFFENYDNSKDTSSNYKRTTGSNVTL-GFPVNENNSVYVGLGHTYKINSFALFYN-- 542
QY 148 SHNVYIMADKQK-NGIKNFKTRHNIEDSGVOLADHYQO-----NTPFGDGPVL 195
DQ 543 -RLNYIQSMKFGKNGIKTN-----DPDFSGWYNYSNLRGYPTFKGVKASLG-GRVT 592
QY 196 LP--DNHYSLTOSALSCKDPNEKXRDHMLVILGFTAAIGITLG 233
DQ 593 IFGSDNKKYKLSADVGQFPLDRDHRVWVSASAGAYANG 632

RESULT 8
SYL HA3IN ID SYL HA3IN STANDARD; PRT; 861 AA.
AC PA3827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
STRAIN=Rd / KW20 / ATCC 51907;
MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fritchman J.L., Glodek A., Kelley J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Cotton M.D.,
RA Weidman J.F., Phillips M.C., Spriggs T., Hedblom E., Brandon R.C.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandt R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
SCIENCE 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32774; AAC22581.1; -

```

CC -1- CATALYTIC ACTIVITY: Release of protein hormones, neuropeptides and
CC bonds from their precursors, generally by cleavage of -Lys-Arg-|-
CC
CC
CC -1- COFACTOR: Calcium.
CC -1- SUBCELLULAR LOCATION: Localized in the secretion granules.
CC -1- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.
CC
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M76705; AAA40945.1; -;
CC EMBL; M83745; AAA41476.1; -;
CC PIR; A41556; KRTCL.
CC HSP; Q45670; IDBI.
CC MEROPS; S08.072; -;
CC InterPro; IPR000209; Peptidase S8.
CC InterPro; IPR002884; Peptidase S8B.
CC InterPro; IPR009020; Protease Inhib.
CC Pfam; PF01483; P_protease; 1.
CC Pfam; PF00082; Peptidase S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC ProDom; PD00717; P domain; 1.
CC PROSITE; PS00136; SUBTILASE ASP; 1.
CC PROSITE; PS00137; SUBTILASE HIS; 1.
CC PROSITE; PS00138; SUBTILASE SER; 1.
CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Calcium; Signal.
FT SIGNAL 1 27
FT CHAIN 28 110 POTENTIAL.
FT CHAIN 111 752 NEUROENDOCRINE CONVERTASE 1.
FT DOMAIN 122 410 CATALYTIC.
FT DOMAIN 739 751 AMPHIPATHIC (POTENTIAL).
FT ACT_SITE 167 167 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 208 208 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 382 382 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 366 366 E -> T (IN REF. 2).
FT CONFLICT 514 514 E -> A (IN REF. 2).
SQ SEQUENCE 752 AA; 84120 MW; F630AD830A076DED CRC64;

Query Match 6.8%; Score 87; DB 1; Length 752;
Best Local Similarity 25.6%; Pred. No. 13;
Matches 53; Conservative 20; Mismatches 70; Indels 64; Gaps 10;

Qy 18 ELDDGVNGHK---RSVSGEGEDATYVKLTCLKFICTTGKLP-----VPWPTLVTLGSLYG 68
Db 540 ERDTSFNGFNDWFMVHTWGENPV--GTWTLKVTDSGRMQNEGRVNWKLILHGT-----594

Qy 69 VQCFSRYPDMKQHDFFKSAMPEGYVQERTIFFKDDGNKTRAEVKFEGDTLVNRIELKG 128
Db 595 ---SSQPEHMKQ-----PRVYT-----SYNTQNDRRGVEKMNVVVEKP 631

Qy 129 IFKEDGNILGHKLEFVYNSHNVYIMADKQKIKVFKIRNIHEDGVSQVLADHYQONT 188
Db 632 TQNSLGNLLVLPK---NSSSSSVEDRRDEQVQAPSAMLR-----LLQSAFSKNT 680

Qy 189 IGDGPVLLPDNHLVSTQSLSKDPNEK 215
Db 681 -----SKQS---SKIPSAK 691

RESULT 10
SYL_XYLEA
ID SYL_XYLEA STANDARD; PRT; 879 AA.
AC Q9PBG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LEURS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarado A.R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kempner E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Martins E.M.F., Matsukuma A.Y.,
RA Marck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quadrio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RL "The genome sequence of the plant pathogen Xylella fastidiosa.";
Nature 406:151-159(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004031; AAF84975.1; ALT_INIT.
DR HAMAP; MF_00049; -; 1.
DR InterPro; IPR002302; Leu-tRNA-synthetase.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009008; ValRS-ilerS_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA tRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99796 MW; 9FDCB992092919E CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 17;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

	Matches	30;	Conservative	18;	Mismatches	39;	Indels	34;	Gaps	5
QY	80	KQHDFFKSAMPEGYVQERTIFFKDDGNNYKTRAEVKF--EGDTLV-----NRIELKG	128							
Db	14	EKHSPIKYYFKFYTKDFKLFASKDKHYRTRAELSFYHENDTLFYAMDFDKSKKKYIIIEY	73							
QY	129	IDFKED-----GNILGHKLEYNTSNHNVIYIMADKQNGIKVNFKIRHNE	173							
Db	74	LQFADESKICAFMRPLLLEYLRQDNKLAEKL-----FGVEFLTTKQE--LSITLLYHKNIE	125							
QY	174	D	174							
Db	126	D	126							

```

RESULT 12
EXPI_ERWCA STANDARD; PRT; 217 AA.
ID EXPI_ERWCA
AC P3882;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Autoinducer synthesis protein expi.
DE EXPI.
GN Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=SCC3193;
RC MEDLINE=93285117; PubMed=8508772;
RA Pirhonen M., Flego D., Heikinheimo R., Palva E.T.;
RT "A small diffusible signal molecule is responsible for the global
RT control of virulence and exoenzyme production in the plant pathogen
RT Erwinia carotovora.";
RT EMBO J. 12:2467-2476(1993).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=SCC3193;
RC Heikinheimo R., Mae A., Flego D., Pirhonen M., Koiv V., Palva E.T.;
RA Submitted (Dec-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR THE SYNTHESIS OF OHHL (N-(3-OXOHEXANOYL)-L-
CC HOMOSERINE LACTONE), AN AUTOINDUCER MOLECULE WHICH BINDS TO EXPR
CC AND THUS ACTS IN VIRULENCE (SOFT ROT DISEASE) THROUGH THE
CC ACTIVATION OF GENES FOR PLANT TISSUE MACERATING ENZYMES.
CC -1- SIMILARITY: Belongs to the autoinducer synthetase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X72891; CAAS1409.1; -.
CC EMBL; X80475; -. NOT ANNOTATED_CDS.
DR PIR; S35324; S35324.
DR InterPro; IPR001690; Autoind synth.
DR Pfam; PF00765; Autoind synth; 1.
DR PRINTS; PR01549; AUTOINDCRSYN.
DR ProDom; PD002752; Autoind synth; 1.
DR PROSITE; PS00949; AUTOINDUCERS SYNTH; 1.
KW Quorum sensing; Autoinducer synthesis.
SQ SEQUENCE 217 AA; 25321 MW; 12F0979D20DFE5D CRC64;

Query Match 6.7%; Score 85.5; DB 1; Length 217;
Best Local Similarity 22.5%; Pred. No. 3.9;
Matches 46; Conservative 27; Mismatches 74; Indels 57; Gaps
4 KGEELFGVVPILVELDGD-----VNGHKFSVSGEGEGDATYGLTLKFICTTGKLP 55
QY | ||||| : | | | : | | : | | : | | : | | : |

```

Db 16 KSELEFT----LRKETFKORLANWAVKINGMEFDQYDDDDNATYLVFGVEGDQVCCSSRLIE 71

QY 56 VPMPTLVTLGSLGYQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKF 115

Db 72 TKYPNMITG-----TFFPY-----FEKIDPEGKYIESRFFVDKARSKTILGNSY 117

QY 116 EGDTL-----VNRIELKGIDFKEDG--NILGHKL-----EYNYNGH 149

Db 118 PVSTMEFLATVNYKSKGY----DGVYTVSHPMILTKRSGWKISIVGQGMSEKHERVY 173

QY 150 NVYIMADKQKNGIKVNFKIRHNIE 173

Db 174 LULFLPVDESQDVLVR-RINHNOE 196

RESULT 13

ITIH3_MESAU STANDARD; PRT; 886 AA.

AC P97280;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Inter-alpha-trypsin inhibitor heavy chain H3 precursor (ITI heavy chain H3) (Inter-alpha-inhibitor heavy chain 3) (HC3).

GN ITIH3.

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.

OC Mesocricetus.

OX NCBI_TaxID=10036;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=97420688; PubMed=9276673; Sinohara H.;

RA Nakatani T., Suzuki Y., Yamamoto T., Sinohara H.;

RT "Molecular cloning and sequencing of cDNAs encoding three heavy-chain precursors of the inter-alpha-trypsin inhibitor in Syrian hamster: implications for the evolution of the inter-alpha-trypsin inhibitor heavy chain family.";

RT J. Biochem. 122:71-82(1997).

RL [2]

RN SEQUENCE OF 31-50; 446-472 AND 504-523, AND SUBUNITS.

RP TISSUE=Plasma;

RX MEDLINE=97018241; PubMed=8864857;

RA Yamamoto T., Yamamoto K., Sinohara H.;

RT "Inter-alpha-trypsin inhibitor and its related proteins in Syrian hamster urine and plasma.";

RL J. Biochem. 120:145-152(1996).

CC -!- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN, INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY SIMILARITY).

CC -!- SUBUNIT: I-alpha-I plasma protease inhibitors are assembled from one or two heavy chains (H1, H2 or H3) and one light chain, bikunin. Inter-alpha-inhibitor (I-alpha-I) is composed of H1, H2 and bikunin, and pre-alpha-like inhibitor (I-alpha-Li) of H2 and bikunin, and pre-alpha-inhibitor (P-alpha-I) of H3 and bikunin.

CC -!- PTM: Heavy chains are interlinked with bikunin via a chondroitin 4-sulfate bridge to the their C-terminal aspartate (By similarity).

CC -!- SIMILARITY: Belongs to the ITIH family.

CC -!- SIMILARITY: Contains 1 VWFA domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; D89287; BAA13940.1; -.

DR InterPro; IPR006587; VIT.

DR InterPro; IPR002035; VWFA_A.

DR Pfam; PF00092; vwa; 1.

DR SMART; SM00609; VII; 1.

DR SMART; SM00327; VWA; 1.

DR PROSITE; PS00234; VWFA; 1.

KW Serine protease inhibitor; Repeat; Signal; Multigene family; Glycoprotein.

FT SIGNAL; 1 18 POTENTIAL.

FT PROPEP; 19 30 BY SIMILARITY.

FT CHAIN; 31 646 INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H3.

FT PROPEP; 647 886 BY SIMILARITY.

FT DOMAIN; 279 439 VWFA.

FT CARBOHYD; 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD; 577 577 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT BINDING; 646 646 CHONDROITIN 4-SULFATE, CROSS-LINK SITE (BY SIMILARITY).

SQ SEQUENCE 886 AA; 99018 MW; AC0594C685257688 CRC64;

Query Match 6.7%; Score 85.5; DB 1; Length 886; Best Local Similarity 23.8%; Pred. No. 21; Mismatches 34; Conservative 36; Indels 19; Gaps 7; Matches 36; Conserved 34; Mismatches 62; Indels 19; Gaps 7;

QY 75 YPDHMKQHDFFKSAMPEGYVQERT-----IFFKDDGNKYKTRAEVKFEGD----TLVNRIE 125

Db 476 YFENAIL-DLTNSYPHFYDGSSETAVAGRLADSDMNFK--ADVKGHALNDLITFTEVD 532

QY 126 LKQID--FKEDGNILGHLEYNVYIMADKQKNGIKVNFKIRHNIEDGVSQVLADHY 183

Db 533 MKEMDAALKEQGVFGNYIERLWAYLTIEQLLEKRN---AHGEKENLTAQALELSKY 589

QY 184 QNTPTIGDGPVLLPDNHYLSTQSALSQDPNE 214

Db 590 HFVTPLTPMVVTKPEDN--EDQTSIADKPGE 618

RESULT 14

SY62_DISOM STANDARD; PRT; 439 AA.

ID SY62_DISOM

AC P24506;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Synaptotagmin B (Synaptic vesicle protein O-p65-B).

GN P65-B.

OS Discopysge ommata (Electric ray).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes; Elasmobranchii; Squalia; Hypnosqualea; Pristiogoralea; Batoidea; Torpediniformes; Narcinoidei; Narcinidae; Discopysge.

OX NCBI_TaxID=7785;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=91273991; PubMed=2054189;

RA Wendland B., Miller K.G., Schilling J., Scheller R.H.;

RT "Differential expression of the p65 gene family.";

RL Neuron 6:993-1007(1991).

CC -!- FUNCTION: May have a regulatory role in the membrane interactions during trafficking of synaptic vesicles at the active zone of the synapse. It binds acidic phospholipids with a specificity that requires the presence of both an acidic head group and a diacyl backbone.

CC -!- SUBUNIT: Homodimer or homotrimer (possible).

CC -!- SUBCELLULAR LOCATION: Synaptic vesicles in neurons.

CC -!- TISSUE SPECIFICITY: Spinal cord, brainstem, midbrain and electric organ.

CC -!- SIMILARITY: Contains 2 C2 domains.

CC -!- SIMILARITY: Belongs to the synaptotagmin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M64276; AAA49228.1; -.
PIR; JH0414; JH0414.
HSSP; P21707; 1BYN.
InterPro; IPR000008; C2.
InterPro; IPR008973; C2_CaLB.
InterPro; IPR002149; LR1.
Pfam; PF00168; C2; 2.
SMART; SM00239; C2; 2.
PROSITE; PS00499; C2_DOMAIN_1; 2.
PROSITE; PS00004; C2_DOMAIN_2; 2.
Transmembrane; Repeat; Synapse; Multigene family; Glycoprotein.
DOMAIN 1 74 VESICULAR (POTENTIAL).
TRANSMEM 75 101 POTENTIAL.
DOMAIN 102 439 CYTOPLASMIC (POTENTIAL).
DOMAIN 153 399 PHOSPHOLIPID BINDING (PROBABLE).
DOMAIN 173 262 C2 DOMAIN 1.
DOMAIN 304 395 C2 DOMAIN 2.
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 46 46
CARBOHYD 46 46
SEQUENCE 439 AA; 49278 MW; 2033F05FD8C69F39 CRC64;

Query Match 6.7%; Score 85; DB 1; Length 439;
Best Local Similarity 19.2%; Pred. No. 10;
Matches 55; Conservative 46; Mismatches 94; Indels 92; Gaps 12;
16 LVELDGVNGHKFSVSGEGSDATYGLTLKFKICTTGKLPV-PWPTLVTLGSLYGV----- 69
38 MPDITDGTNEATGVPGEKND-VFEKLEKFMNELQKIFLPWALIAIVSGILLTLC 96
70 -----OCFSKYPDHMKQHDFFKSAPEGYVQERTIFPKDDGNKYKRAEVKFEK----- 117
97 CLCICKCKCKCKKCKKCKGK-----KNDINWK-----DVKSGSGNQDD 138
118 --DTLVNRIELKIDFKEDGNI--LGHKLEYNVNH----- 149
139 DAETGLTEGEDKEEKEEKLKIQSLFYDFOANQLTVGIIQAELPALDMGSGSDPY 198
150 -NVYIMADKKN-GIKVN-----FKIRHIEDGSVOLA-----DHYQQNTPI 189
199 VKVFLPPDKKKYETKQKTLNFTFNESVFVKPYQELGKTLMAVYDFDRFSKHDCI 258
190 GDGVPVLLPD-----NHYLSQSAISKOPNEKRDHMLVLLGFVTAAG 229
259 QGVTVLMTKVDLGQQLBEWRDLSEAEKEPEKLGDICTSLRYVPTAG 305

RESULT 15
ID GLNA AZOCA STANDARD; PRT; 468 AA.
AC P94125;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
GN GLNA.
OS Azorhizobium caulinodans.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Hyphomicrobiaceae; Azorhizobium.
OX NCBI_TaxID=7;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORS571;
RX MEDLINE=97315229; PubMed=9171403;
RA Michel-Reydellet N., Desnoues N., Elmerich C., Kaminski P.A.;
RT "Characterization of Azorhizobium caulinodans glnB and glnA genes:
involvement of the P(II) protein in symbiotic nitrogen fixation.";
RL J. Bacteriol. 179:3580-3587(1997).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +

L-glutamine.
-!- ENZYME REGULATION: The activity of this enzyme is controlled by adenylation. The fully adenylation enzyme complex is inactive.
-!- SUBUNIT: Oligomer of 12 subunits arranged in the form of two hexagons (By similarity).
-!- SUBCELLULAR LOCATION: Cytoplasmic.
-!- SIMILARITY: Belongs to the glutamine synthetase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Y10213; CAA71265.1; -.
HSSP; P06201; ILGR.
InterPro; IPR008147; Gln_synt_beta.
InterPro; IPR008146; Gln_synt_C.
InterPro; IPR004809; GlnA.
InterPro; IPR001637; GlnA_adenyltn_s.
Pfam; PF00120; gln-synt; 1.
Pfam; PF03951; gln-synt_N; 1.
ProDom; PD001057; Gln_synt_C; 1.
TIGRFAM; TIGR00653; GlnA; 1.
PROSITE; PS00180; GLNA_ATP; 1.
PROSITE; PS00181; GLNA_ATP; 1.
PROSITE; PS00182; GLNA_ADENYLATION; 1.
KW Nitrogen fixation; Ligase.
FT BINDING 397 397 AMP (UNDER CONDITIONS OF ABUNDANT GLUTAMINE) (BY SIMILARITY).
SQ SEQUENCE 468 AA; 52359 MW; 36C759B9C95DS303 CRC64;

Query Match 6.6%; Score 84.5; DB 1; Length 468;
Best Local Similarity 24.9%; Pred. No. 12;
Matches 45; Conservative 21; Mismatches 60; Indels 55; Gaps 10;
37 DATYGLKLTIKFICTTGKLPVWPPTLVTLGSLYGVQCFSPYDPDHMKQHDFFKSAPEGYVQ 96
78 DPFFSETTLISVVC-----DVLEPT--TGEFYG-----RDP-----RGIKKAM--AYLOS 118
97 R-----TIFFKDDGNKYKRAEVKFEKDTLVNRIELKIDFKEDGNI-----LGHKLEYN 146
119 TGIGDTVFFGPEAEFFIDDFKADPYNTGFKLDSIELPTNGDIDYEGNGLHRIKTKG 178
147 NSHNV-----YIMADKKNKGIKYNFKIRHIEDGSVQ-----LADHY 183
179 GYFVPVPLDSAQDMRSEMLASMAKMGAKVE---KHHEVASAQHGLGLKFGQLVTMADHL 235
184 Q 184
236 Q 236

Search completed: June 21, 2004, 15:55:19
Job time : 8.55556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds
(without alignments)
2458.984 Million cell updates/sec

Title: US-09-887-784-64G
Perfect score: 1275
Sequence: 1 MVSKEBELTGVVPIVLVELD.....VLIGFVTAAGITLGMDELYK 239
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1236	96.9	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1233	96.7	238	5 Q93125	Q93125 aequorea vi
3	1231	96.5	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1230	96.5	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1198	94.0	238	5 Q17105	Q17105 aequorea vi
6	1183	92.8	238	5 Q17106	Q17106 aequorea vi
7	1078	84.5	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1074	84.2	238	5 Q8WP95	Q8wp95 aequorea ma
9	1070	83.9	238	5 Q8WTC4	Q8wtc4 aequorea ma
10	1068	83.8	238	5 Q8WTC5	Q8wtc5 aequorea ma
11	1067	83.7	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1067	83.7	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1065	83.5	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1063	83.4	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5 Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5 Q7Z0W5	Q7z0w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	244.5	19.2	236	5	Q8TGU0	Q8tgu0 dendronephth
19	240	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyl
20	239.5	18.8	225	5	Q7Z0W9	Q7z0w9 montastraea
21	238.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q7Z0W4	Q7z0w4 montastraea
23	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
24	229	18.0	225	5	Q8T5F1	Q8t5f1 montastraea
25	212.5	16.7	259	5	Q8MMA2	Q8mma2 agaricia fr
26	212	16.6	239	5	Q8VMA1	Q8vma1 agaricia ag
27	211	16.5	227	5	Q7Z0W6	Q7z0w6 montastraea
28	211	16.5	234	5	Q7Z0W7	Q7z0w7 montastraea
29	206	16.2	234	5	Q8T5F2	Q8t5f2 montastraea
30	206	16.2	234	5	Q8MU47	Q8mu47 montastraea
31	205.5	16.1	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	204.5	16.0	232	5	Q9GP15	Q9gpi5 anemonia su
33	203.5	16.0	238	5	Q9BLV9	Q9blv9 renilla mue
34	203	15.9	221	5	Q9SP04	Q9sp04 gonloporo t
35	203	15.9	227	5	Q962P9	Q962p9 montastraea
36	203	15.9	227	5	Q7Z0W8	Q7z0w8 montastraea
37	200.5	15.7	225	5	Q9U6Y8	Q9u6y8 discosoma s
38	200.5	15.7	232	5	Q9GZ28	Q9gz28 anemonia s
39	199	15.6	227	5	Q95VT0	Q95vt0 montastraea
40	198.5	15.6	222	5	Q7Z168	Q7z168 cerianthus
41	198.5	15.6	225	5	Q8TG19	Q8tg19 radianthus
42	198	15.5	235	5	Q8T5F0	Q8t5f0 scolymia cu
43	197.5	15.5	214	5	Q86LV7	Q86lv7 meandrina m
44	196.5	15.4	214	5	Q86LV8	Q86lv8 meandrina m
45	196.5	15.4	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	ID	Q8GHE2	PRELIMINARY;	PRT;	238 AA.
AC	Q8GHE2;				
DT	01-MAR-2003 (Tremblrel. 23, Created)				
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)				
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)				
DE	Green fluorescence protein.				
GN	2289GFP.				
OS	Azotobacter vinelandii.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Azotobacter.				
OX	NCBI_TaxID=354;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DSM2289;				
RA	Koranyi P., Berenyl M., Burg K.;				
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria				
RL	Azomonas and Azotobacter.";				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF324408; AAN86140.1; -				
DR	GO; GO:0006091; P:energy pathways; IEA.				
DR	InterPro; IPR0009017; GFP like.				
DR	InterPro; IPR000786; Green_fl_protein.				
DR	Pfam; PF01353; GFP; 1.				
DR	PRINTS; PR01229; GFP; 1.				
DR	ProDom; PD013756; Green_fl_protein; 1.				
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;				

Query Match	96.9%;	Score 1236;	DB 2;	Length 238;
Best Local Similarity	97.5%;	Pred. No. 4.9e-96;		
Matches 232;	Conservative	2;	Mismatches 4;	Indels 0; Gaps 0;
QY	2	VSKGELFTGVVPIVLVELDGVNGHKFSVSGEGDATYKGLTKFKICTTCKLPVPMPTL	61	
Db	1	MSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYKGLTKFKICTTCKLPVPMPTL	60	
QY	62	VTGLSYGVQCFRYPDHMKQHDFFKSPMEGVYQERTIPFKDGNKYKTRAEVKEGDTLV	121	

Db 61 VTTFSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

Qy 122 NRIELKGIDPKEDGNILGHKLEYNHSHVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 181

Db 121 NRIELKGIDPKEDGNILGHKLEYNHSHVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 180

Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMHVLGFTVTAAGITLGMDELYK 239

Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMHVLGFTVTAAGITLGMDELYK 238

RESULT 2

Q93125 PRELIMINARY; PRT; 238 AA.

AC Q93125; (T-EMBLrel. 02, Created)

DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)

DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescent protein mutant 3.

GN GFP.

OS Aequorea victoria (Jellyfish).

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=6100;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96305137; PubMed=8707053;

RA Cormack B.P., Valdivia R.H., Falkow S.;

RA "FACS-optimized mutants of the green fluorescent protein (GFP).";

RL Gene 173:33-38(1996).

RN [2]

RP SEQUENCE FROM N.A.

RA Cormack B.P., Bertam G., Egerton M., Gow N.A.R., Falkow S.,

RA Brown A.J.P.;

RA "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene

RT expression in Candida albicans.;"

RL Microbiology 0:0-0(1996).

DR EMBL; U73901; AAB18957.1; -.

DR HSP; P42212; IAFP.

DR GO; CC:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP.

DR PRODOR; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 96.7%; Score 1233; DB 5; Length 238;

Best Local Similarity 97.1%; Pred. No. 8.7e-96;

Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 VSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDGYGKLTGKLTCTTGKLPVWPPTL 61

Db 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDGYGKLTGKLTCTTGKLPVWPPTL 60

Qy 62 VTGLSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121

Db 61 VTTFSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

Qy 122 NRIELKGIDPKEDGNILGHKLEYNHSHVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 181

Db 121 NRIELKGIDPKEDGNILGHKLEYNHSHVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 180

Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMHVLGFTVTAAGITLGMDELYK 239

Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMHVLGFTVTAAGITLGMDELYK 238

RESULT 3

Q93125 PRELIMINARY; PRT; 238 AA.

AC Q93125; (T-EMBLrel. 02, Created)

DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)

DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescent protein mutant 3.

GN GFP.

OS Aequorea victoria (Jellyfish).

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=6100;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96305137; PubMed=8707053;

RA Cormack B.P., Valdivia R.H., Falkow S.;

RA "FACS-optimized mutants of the green fluorescent protein (GFP).";

RL Gene 173:33-38(1996).

RN [2]

RP SEQUENCE FROM N.A.

RA Cormack B.P., Bertam G., Egerton M., Gow N.A.R., Falkow S.,

RA Brown A.J.P.;

RA "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene

RT expression in Candida albicans.;"

RL Microbiology 0:0-0(1996).

DR EMBL; U73901; AAB18957.1; -.

DR HSP; P42212; IAFP.

DR GO; CC:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP.

DR PRODOR; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 96.7%; Score 1233; DB 5; Length 238;

Best Local Similarity 97.1%; Pred. No. 8.7e-96;

Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 VSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDGYGKLTGKLTCTTGKLPVWPPTL 61

Db 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDGYGKLTGKLTCTTGKLPVWPPTL 60

Qy 62 VTGLSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121

Db 61 VTTFSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

Qy 122 NRIELKGIDPKEDGNILGHKLEYNHSHVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 181

Db 121 NRIELKGIDPKEDGNILGHKLEYNHSHVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 180

Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMHVLGFTVTAAGITLGMDELYK 239

Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMHVLGFTVTAAGITLGMDELYK 238

RESULT 4

Q93125 PRELIMINARY; PRT; 238 AA.

AC Q93125; (T-EMBLrel. 02, Created)

DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)

DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Azotobacter vinelandii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Azotobacter.

OX NCBI_TaxID=354;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96305137; PubMed=8707053;

RA Koranyi P., Berenyi M., Burg K.;

RA "Occurrence of green fluorescence protein in diazotrophic bacteria

RT Azomonas and Azotobacter.;"

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF324406; AAN86137.1; -.

DR GO; CC:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP.

DR PRODOR; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 26840 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.5%; Score 1230; DB 2; Length 238;

Best Local Similarity 97.1%; Pred. No. 1.3e-95;

Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 VSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDGYGKLTGKLTCTTGKLPVWPPTL 61

Db 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDGYGKLTGKLTCTTGKLPVWPPTL 60

Qy 62 VTGLSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121

Db 61 VTTFSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

Qy 122 NRIELKGIDPKEDGNILGHKLEYNHSHVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 181

Db 121 NRIELKGIDPKEDGNILGHKLEYNHSHVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 180

Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMHVLGFTVTAAGITLGMDELYK 239

Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDMHVLGFTVTAAGITLGMDELYK 238

Best Local Similarity 97.1%; Pred. No. 1.6e-95; Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;									
QY	2	VSKGEELFTGVVPILVELDGVN	GKHFVS	GE	GDATY	GKLT	FKICT	GTGKLP	VPWPPTL 61
Db	1	MSKGEELFTGVVPILVELDGVN	GKHFVS	GE	GDATY	GKLT	FKICT	GTGKLP	VPWPPTL 60
QY	62	VTGLSYGVQCFSRYPDHMKQHD	FFKS	AM	PEGVYQERT	IFFK	DDGN	YKTRAE	VKFE
Db	61	VTTFSGYGVQCFSRYPDHMKR	HDFF	FKS	AM	PEGVYQERT	IFFK	DDGN	YKTRAE
QY	122	NRIELKGIDFKEDGNILGH	LEYN	SHN	YIMAD	KQNGIK	VNF	KIRN	IEDG
Db	121	NRIELKGIDFKEDGNILGH	LEYN	SHN	YIMAD	KQNGIK	VNF	KIRN	IEDG
QY	182	HYQONTPIGDGPVLLPD	NHLS	TQS	ALS	KDPNE	KRDH	MLLV	FGVTA
Db	181	HYQONTPIGDGPVLLPD	NHLS	TQS	ALS	KDPNE	KRDH	MLLV	FGVTA
RESULT 5									
ID	Q17105	PRELIMINARY; PRT; 238 AA.							
AC	Q17105								
DT	01-NOV-1996	(TrEMBLrel. 01, Created)							
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)							
DE	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)							
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83960; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								
OX	Aequoreidae; Aequorea.								
RN	[1]	TaxID=6100;							
RP	SEQUENCE FROM N.A.								
RA	Watkins J.N., Campbell A.K.;								
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; X83959; CAA58789.1; -								
DE	Green fluorescent protein (Fragment).								
GN	GFP.								
OS	Aequorea victoria	(Jellyfish).							
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;								


```
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL33912.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match      83.7%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 8.1e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGWPIVLVELDGVNKHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFKICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLGYGIOCFARYPEHMKMNDFFKSAMPEGYVQERTIFFQDDGKYKTRGEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNHNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 NRIELKGMDPKEDGNILGHKLEYNFNHNSHNYYIMPDKANNGKLVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 HYQTNVPLGDGPVLLIPINHYLSFQTALSKDRNETRDHMLVLEFFSACGHTGHMDELYK 238

RESULT 12
Q8WTC9 PRELIMINARY; PRT; 238 AA.
ID Q8WTC9
AC Q8WTC9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL33913.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match      83.7%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 8.1e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGWPIVLVELDGVNKHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFKICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLGYGIOCFARYPEHMKMNDFFKSAMPEGYVQERTIFFQDDGKYKTRGEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNHNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 NRIELKGMDPKEDGNILGHKLEYNFNHNSHNYYIMPDKANNGKLVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 HYQTNVPLGDGPVLLIPINHYLSFQTALSKDRNETRDHMLVLEFFSACGHTGHMDELYK 238

RESULT 13
Q8WTC7 PRELIMINARY; PRT; 238 AA.
ID Q8WTC7
AC Q8WTC7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL33912.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match      83.7%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 8.1e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGWPIVLVELDGVNKHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFKICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLGYGIOCFARYPEHMKMNDFFKSAMPEGYVQERTIFFQDDGKYKTRGEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNHNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 NRIELKGMDPKEDGNILGHKLEYNFNHNSHNYYIMPDKANNGKLVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 HYQTNVPLGDGPVLLIPINHYLSFQTALSKDRNETRDHMLVLEFFSACGHTGHMDELYK 238

RESULT 14
Q8WTC8 PRELIMINARY; PRT; 238 AA.
ID Q8WTC8
AC Q8WTC8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435429; AAL33914.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match      83.7%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 8.1e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGWPIVLVELDGVNKHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFKICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLGYGIOCFARYPEHMKMNDFFKSAMPEGYVQERTIFFQDDGKYKTRGEVKEGDTLV 120
```


Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1268	99.6	239	5	AAE17518	Enhanced
2	1268	99.6	363	6	ABR40352	Human ami
3	1268	99.6	893	4	AAAG5781	Amino aci
4	1268	99.6	1132	4	AAAG65782	Amino aci
5	1260	99.0	239	5	AAE17517	Enhanced
6	1257	98.7	239	3	AAAB22882	Enhanced
7	1257	98.7	239	3	AAAY54349	Amino aci
8	1257	98.7	239	3	AAAY79584	EGFP sign
9	1257	98.7	239	4	AAAB50804	Jellyfish
10	1257	98.7	239	4	AAAB85900	A. victor
11	1257	98.7	239	4	AAAB31171	Amino aci
12	1257	98.7	239	5	AAAG66198	A. victor
13	1257	98.7	239	5	ABG94444	Protease
14	1257	98.7	239	5	AAE14539	Aequorea
15	1257	98.7	239	6	AAE34958	Aequorea
16	1257	98.7	239	6	AAAG79829	Green flu
17	1257	98.7	239	6	ABR83616	Green flu
18	1257	98.7	239	6	ADA38074	Aequorea
19	1257	98.7	239	7	ABU63204	Aequorea
20	1257	98.7	239	7	ADC18358	EGFP (erh)
21	1257	98.7	239	7	ABW00914	Aequorea
22	1257	98.7	239	7	ADE28570	Enhanced
23	1257	98.7	246	7	ABM79011	Enhanced
24	1257	98.7	248	5	AAAG68319	Jellyfish
25	1257	98.7	259	5	AAU99804	Biomembra

XX The invention relates to a fluorescent protein derived from green
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
 CC at F64L and E222G has a bigger compared to other GFP's making it very
 CC suitable for high throughput screening due to better resolution. The
 CC fluorescent protein is useful in *in vitro* assays for measuring protein
 CC kinase activity or dephosphorylation activity, or for measuring protein
 CC redistribution. The fluorescent protein is useful in studying cellular
 CC functions in living cells; as protein tags in transgenic animals, living
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.
 CC The fluorescent protein is also useful as a cell or organelle integrity
 CC marker, a marker for changes in cell morphology, as transfection marker,
 CC and as a marker to be used in combination with fluorescence activated
 CC cell sorting (FACS). The novel proteins can also be used as reporters to
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
 CC protein is also useful as markers in transcriptional and translational
 CC fusions for performing transposon vector mutagenesis and as a reporter
 CC for bacterial detection. Transposons encoding the fluorescent protein are
 CC useful for screening promoters and for tagging plasmids and chromosomes.
 CC The fluorescent protein engineered into the genome of a phage is useful
 CC for designing diagnostic tool. The present sequence is a DNA encoding
 CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
 XX
 SQ Sequence 239 AA;

Query Match 99.6%; Score 1268; DB 5; Length 239;
 Best Local Similarity 99.8%; Pred. No. 2e-122;
 Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKGLTKLFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKGLTKLFICTTGKLPVPWPT 60

QY 51 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
 DB 51 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 131 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 DB 131 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 2
 ABR40352
 ID ABR40352 standard; protein; 363 AA.
 AC ABR40352;
 XX
 DT 08-JUL-2003 (first entry)
 DE Human amino acid sequence SEQ ID NO: 6.
 DE Human; heterologous conjugate; intracellular protein.
 KW Homo sapiens.
 OS Aequoria victoria.
 OS WO2003029827-A2.
 PN
 XX 10-APR-2003.
 PD
 XX 01-OCT-2002; 2002WO-DK000651.
 XX 01-OCT-2001; 2001DK-00001433.
 PR 11-OCT-2001; 2001US-0328896P.
 XX (BIOL-) BIOMAGE AS.
 PA Terry BR, Nielsen SJ;
 XX

XX WPI; 2003-430211/40.
 DR N-PSDB; ACC72604.
 XX
 PT Novel cell for identifying modulators of protein interaction, contains a
 PT first conjugate comprising anchor protein, second conjugate having type B
 PT interactor protein and third conjugate with detectable group.
 XX
 PS Disclosure; Page 112-113; 118pp; English.
 XX
 CC The invention relates to a novel cell, comprising three heterologous
 CC conjugates (HC), a first HC (HC1) comprising an anchor protein that
 CC specifically binds to an internal structure within the cell conjugated to
 CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
 CC type B conjugated to a first protein of interest, and a third HC (HC3)
 CC comprising a second protein of interest conjugated to detectable group.
 CC The cell is useful for detecting if a compound disrupts or induces the
 CC interaction between two intracellular proteins. The cell is also useful
 CC for screening compounds that modulate the interaction between two
 CC intracellular proteins. The present sequence is used in the
 CC exemplification of the invention
 XX
 SQ Sequence 363 AA;

Query Match 99.6%; Score 1268; DB 6; Length 363;
 Best Local Similarity 99.6%; Pred. No. 3.7e-122;
 Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKGLTKLFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKGLTKLFICTTGKLPVPWPT 60

QY 61 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
 DB 61 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3
 AAG65781
 ID AAG65781 standard; protein; 893 AA.
 XX
 AC AAG65781;
 XX
 DT 07-JAN-2002 (first entry)
 DE Amino acid sequence of HSPDE4A1-E222G fusion protein.
 DE PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
 KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
 KW fusion protein.
 XX Homo sapiens.
 OS Aequorea victoria.
 OS WO200179526-A2.
 PN
 XX 25-OCT-2001.
 PD
 XX 11-APR-2001; 2001WO-DK000264.
 XX 17-APR-2000; 2000DK-00000651.
 PR 29-MAY-2000; 2000DK-00000849.
 XX (BIOI-) BIOIMAGE AS.
 PA
 XX


```
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
XX
DR WPI: 2001-611727/70.
DR N-PSDB; AA166852.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 156-160; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
XX protein
XX
XX Sequence 893 AA;
XX
XX Query Match 99.6%; Score 1268; DB 4; Length 893;
XX Best Local Similarity 99.6%; Pred. No. 1.4e-121;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 714
XX
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
XX
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
XX
QY 181 DHYQONTPIGDGPFVLLPDNHHYLSQTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPFVLLPDNHHYLSQTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 893
XX
RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA.
XX
AC AAG65782;
XX
DT 07-JAN-2002 (first entry)
XX
XX Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX fusion protein.
XX
XX Homo sapiens.
XX Aequorea victoria.
XX
XX WO200179526-A2.
XX
```

```
PD 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-DK000264.
XX
XX 17-APR-2000; 2000DK-00000651.
XX 29-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
XX Praestegaard M;
XX
XX WPI: 2001-611727/70.
XX N-PSDB; AA166853.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
XX compounds for treating CNS and inflammatory disease comprises identifying
XX compounds which remove PDE4 spots.
XX
XX Example 1; Page 162-167; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
XX protein
XX
XX Sequence 1132 AA;
XX
XX Query Match 99.6%; Score 1268; DB 4; Length 1132;
XX Best Local Similarity 99.6%; Pred. No. 2e-121;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 953
XX
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
XX
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
XX
QY 181 DHYQONTPIGDGPFVLLPDNHHYLSQTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPFVLLPDNHHYLSQTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 1132
XX
RESULT 5
AAE17517
ID AAE17517 standard; protein; 239 AA.
XX
AC AAE17517;
XX
DT 22-APR-2002 (first entry)
XX
XX Enhanced F64L jellyfish green fluorescent protein mutant.
XX
```

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
 KW cellular function; genetic reporter; mutant; Stoke's shift; mtein.
 XX
 OS Aequorea victoria.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 65
 FT /note= "Wild type Phe substituted with Leu; This
 FT corresponds to position 64 in the wild type protein"
 XX
 XX WO200198338-A2.
 XX
 XX 27-DEC-2001.
 XX
 XX 18-JUN-2001; 2001WO-EP006848.
 XX
 XX 19-JUN-2000; 2000DK-00000953.
 PR 20-JUN-2000; 2000US-0212681P.
 PR 10-MAY-2001; 2001DK-00000739.
 PR 10-MAY-2001; 2001US-0290170P.
 XX
 XX (BIOI-) BIOIMAGE AS.
 PA
 XX Bjorn .3P, Pagliaro L, Thastrup O;
 XX
 XX WPI; 2002-098224/13.
 DR N-PSDB; AAD28162.
 XX
 XX Novel fluorescent protein in in vitro assay for measuring protein kinase
 PT activity or dephosphorylation activity, or for measuring protein
 PT redistribution, has a green fluorescent protein with F64L and E222G
 PT mutation.
 XX
 XX Example 1; Page 35; 41pp; English.
 XX
 XX The invention relates to a fluorescent protein derived from green
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
 CC at F64L and E222G has a bigger compared to other GFP's making it very
 CC suitable for high throughput screening due to better resolution. The
 CC fluorescent protein is useful in invitro assays for measuring protein
 CC kinase activity or dephosphorylation activity, or for measuring protein
 CC redistribution. The fluorescent protein is useful in studying cellular
 CC functions in living cells; as protein tags in transgenic animals, living
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.
 CC The fluorescent protein is also useful as a cell or organelle integrity
 CC marker, a marker for changes in cell morphology, as transfection marker,
 CC and as a marker to be used in combination with fluorescence activated
 CC cell sorting (FACS). The novel proteins can also be used as reporters to
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
 CC protein is also useful as markers in transcriptional and translational
 CC fusions for performing transposon vector mutagenesis and as a reporter
 CC for bacterial detection. Transposons encoding the fluorescent protein are
 CC useful for screening promoters and for tagging plasmids and chromosomes.
 CC The fluorescent protein engineered into the genome of a phage is useful
 CC for designing diagnostic tool. The present sequence is enhanced F64L
 CC jellyfish green fluorescent protein (GFP) mutant
 XX
 XX Sequence 239 AA;
 SQ
 Query Match 99.0%; Score 1260; DB 5; Length 239;
 Best Local Similarity 99.2%; Pred. No. 1.3e-121;
 Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MVSGBELFTGVVPIVLVDGDNVNGHKFSVSGEGGDATYVKLTLPKICTTGKLPVWPWT 60
 DB 1 MVSGBELFTGVVPIVLVDGDNVNGHKFSVSGEGGDATYVKLTLPKICTTGKLPVWPWT 60
 QY 51 LVTLTSGVQCFGRYPDPHMQHDFFKSAMPEGVGVQERTIFFKDDGNKTRAEVKFGDYL 120
 DB 51 LVTLTSGVQCFGRYPDPHMQHDFFKSAMPEGVGVQERTIFFKDDGNKTRAEVKFGDYL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNVNSHNVIADKQNGIKVNFKIRHNIEDSGVQLA 180

DB 121 VNRIELKGIDFKEDGNILGHKLEYNVNSHNVIADKQNGIKVNFKIRHNIEDSGVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
 RESULT 6
 AAB22882
 ID AAB22882 standard; protein; 239 AA.
 XX
 XX AAB22882;
 XX
 XX 10-JAN-2001 (first entry)
 DT
 XX
 XX Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.
 DE
 XX
 XX Bioreactor protein; fusion protein; recognition site;
 KW cellular targeting sequence; cellular localisation; fluorescent protein;
 KW protease activity detection; toxin detection; cellular stress detection;
 KW drug discovery; cell based screening.
 XX
 XX Aequorea victoria.
 OS Synthetic.
 OS
 XX WO200050872-A2.
 PN
 XX 31-AUG-2000.
 PD
 XX
 XX 25-FEB-2000; 2000WO-US004794.
 PF
 XX
 XX 26-FEB-1999; 99US-0122152P.
 PR 08-MAR-1999; 99US-0123399P.
 PR 12-JUL-1999; 99US-00352171.
 XX
 XX (CELL-) CELLOMICS INC.
 PA
 XX Giuliano KA, Kapur R;
 PI
 XX WPI; 2000-594086/56.
 DR N-PSDB; AAA93373.
 XX
 XX Automated cell-based characterization of toxin by contacting cells
 PT containing luminescent reporter molecules with test substance and
 PT analyzing optically.
 XX
 XX Example 11; Fig 29A; 336pp; English.
 PS
 XX The invention relates to systems, methods and reagents for cell-based
 CC screening or detection of compounds which affect particular biological
 CC functions. The methods of the invention utilise fluorescent bioreactor
 CC molecules which, when acted on by a compound of interest, cause an
 CC alteration in the cellular distribution of at least the fluorescent
 CC moiety. In one embodiment, the biosensors comprise heat shock proteins
 CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent
 CC protein (GFP), or derivatives thereof). Such biosensors are located in
 CC the cytoplasm, but on stress activation translocate to the nucleus. In
 CC another embodiment bioreactor proteins can be used to detect protease
 CC activity. Such protease bioreactor fusion proteins comprise one or more
 CC fluorescent proteins; a recognition signal which is cleaved by the
 CC protease; and at least one cellular localisation signal. The latter two
 CC components may be components of a single protein which is acted upon by
 CC the protease, or may be from heterologous sources. Due to the
 CC localisation signal, the bioreactor protein is localised to a particular
 CC region of the cell. Once acted on by the protease of interest, the
 CC fluorescent protein is cleaved from the localisation sequence, and is
 CC free to migrate to other locations within the cell. The presence of a
 CC second localisation signal attached to the fluorescent protein enables
 CC the fluorescent protein to be directed to a different cellular
 CC compartment after cleavage of the protease recognition sequence. The
 CC change in distribution of the fluorescent protein can be detected using
 CC imaging methods with a high degree of spatial resolution. The methods and

CC biosensors of the invention can be used to investigate a wide range of
CC cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased quantity
CC and speed of data collection and faster evaluation of drug candidates.
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used
CC as components of biosensor fusion proteins of the invention
XX
SQ Sequence 239 AA;

Query Match 98.7%; Score 1257; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTLKFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 7
AAV54349
ID AAY54349 standard; protein; 239 AA.
XX AAY54349;
XX
XX
XX 06-APR-2000 (first entry)
XX
DE Amino acid sequence of the mutant green fluorescent protein EGFP.
XX
XX Fluorescent protein; green fluorescent protein; emission intensity;
KW fluorescence; pH detection; pH sensor; EGFP.
XX
XX Synthetic.
OS Aequorea victoria.
XX
FH Key Location/Qualifiers
FT Misc-difference 65 /note= "wild type Phe substituted with Leu"
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"
FT Misc-difference 232 /note= "wild type His substituted with Leu"
FT
XX
PN W09964592-A2.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US012850.
XX
PR 09-JUN-1998; 98US-00094359.
PR 13-OCT-1998; 98US-00172063.
XX
XX (REG) UNIV CALIFORNIA.
PA (UYOR-) UNIV OREGON STATE.
XX

PI Tsien RY, Llopis J, Wachtler RM;
XX
DR WPI; 2000-116540/10.
DR N-PSDB; AA245642.
XX
PT New functional engineered green fluorescent proteins, used for measuring
PT the pH in biological samples and cells.
XX
PS Disclosure; Page 9; 89pp; English.
XX
CC The present sequence represents a functional engineered fluorescent
CC protein based on the Aequorea green fluorescent protein (GFP). The
CC emission intensity changes as pH varies between 5 and 10 of the present
CC protein are novel. The functional engineered fluorescent proteins show
CC reversible changes in fluorescence over physiological pH ranges. They can
CC be used for determining the pH of samples and cells. The polynucleotides
CC can also be used to produce transgenic animals. The fluorescent protein
CC pH sensors can be delivered to cells in the form of polynucleotides
CC encoding the protein sensor fused to a targeting signal. The targeting
CC signal directs the expression of the protein sensors to restricted cell
CC locations. This makes it possible to measure the pH of a precisely
CC defined cellular region or organelle
XX
SQ Sequence 239 AA;

Query Match 98.7%; Score 1257; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTLKFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 8
AAV79584
ID AAY79584 standard; peptide; 239 AA.
XX
XX AAY79584;
XX
XX 29-AUG-2000 (first entry)
XX
DE EGFP signal domain.
XX
KW Protease; biosensor; EGFP; signal peptide; cell screening; assay;
KW analysis; drug discovery.
XX
OS Unidentified.
XX
PN W0200026408-A2.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99WO-US025431.
XX
PR 30-OCT-1998; 98US-0106308P.
PR 26-MAY-1999; 99US-0136078P.
XX
XX (CELL-) CELLOMICS INC.
XX
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX WPI; 2000-365644/31.
DR N-PSDB: AAA27573.
XX
XX Recombinant nucleic acid encoding a protease biosensor useful for
PT fluorescence based cell and molecular biochemical assays for drug
PT discovery comprising three operably linked nucleic acid sequences.
XX
XX Claim 14; Fig 29A; 218pp; English.
XX
XX The present sequence is that of the EGFP signal domain, which can be
CC included in novel recombinant protease biosensors (PBs) of the invention.
CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)
CC comprising at least 1 detectable polypeptide signal such as the present
CC sequence; a second domain (see AAY79588-622) comprising at least 1
CC protease recognition site; and a third domain (see AAY79623-37)
CC comprising at least 1 reactant target sequence. A recombinant nucleic
CC acid (see AAA27627-43) encoding the PB, an expression vector, and a
CC genetically engineered host cell are also claimed. A claimed method for
CC identifying compounds that modify protease activity in a cell involves
CC contacting a host cell that possesses the recombinant PB with a test
CC compound, and determining the PB distribution in the host cell, where
CC changes in the distribution of the PB are correlated with modification of
CC protease activity by the test compound. Claimed kits for identifying
CC compounds that modify protease activity in a host cell include the
CC recombinant nucleic acid, or the recombinant PB, or the vector, or the
CC host cell. The PB is useful in high content screens to detect in vivo
CC activation of enzymatic activity, and to identify specific activity based
CC on cleavage of a known recognition motif
XX
XX Sequence 239 AA;
SQ

Query Match 98.7%; Score 1257; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60
Db |||||
1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60
QY 51 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db |||||
51 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db |||||
121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 131 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db |||||
131 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 9
AAB50804
ID AAB50804 standard; protein; 239 AA.
XX
AC AAB50804;
XX
DT 14-MAR-2001 (first entry)
XX
DE Jellyfish GFP mutant EGFP.
XX
KW Aequorea victoria; jellyfish; fluorescent protein indicator;
KW green fluorescent protein; GFP; linker moiety; sensor;
KW calmodulin-binding domain; mutant; mutein.
XX
OS Aequorea victoria.
XX
PN WO200071565-A2.
XX
PD 30-NOV-2000.
XX

PF 17-MAY-2000; 2000WO-US013684.
XX
PR 21-MAY-1999; 99US-00316919.
PR 21-MAY-1999; 99US-00316920.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Tsien RY, Baird GA;
XX
DR WPI; 2001-032017/04.
DR N-PSDB; AAC90488.
XX
XX Novel fluorescent proteins comprising a sensor protein inserted into
PT them, useful for measuring the response of a sensor biological, chemical,
PT electrical or physiological parameter in vivo or in vitro.
XX
XX Disclosure; Page 24; 94pp; English.
XX
XX The present sequence is a fluorescent protein used in the construction of
CC a fluorescent protein indicator. The indicator comprises a sensor
CC polypeptide that is responsive to a chemical, biological, electrical or
CC physiological parameter, and a fluorescence protein functional group. The
CC sensor polypeptide is operatively inserted into the fluorescent moiety.
CC The fluorescent indicator is useful for detecting the presence of a
CC response inducing member in a sample. The method involves contacting the
CC sample with the indicator and detecting a change in fluorescence, in
CC which a change is indicative of the effect of the parameter on the sensor
CC polypeptide. The novel fluorescent proteins are advantageous due to their
CC reduced size as compared to the FRET (fluorescence resonance energy
CC transfer)-based sensors
XX
XX Sequence 239 AA;
SQ

Query Match 98.7%; Score 1257; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60
Db |||||
1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db |||||
61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db |||||
121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db |||||
181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 10
AAB85900
ID AAB85900 standard; protein; 239 AA.
XX
AC AAB85900;
XX
DT 30-NOV-2001 (first entry)
XX
DE A. victoria green fluorescent protein (GFP) and linker sequence.
XX
KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;
KW green fluorescent protein.
XX
OS Synthetic.
OS Aequorea victoria.
XX
PN WO200168706-A1.
XX

PD 20-SEP-2001.
XX
XX
PF 14-MAR-2001; 2001WO-US008071.
XX
PR 15-MAR-2000; 2000US-0189698P.
XX
XX (MERI) MERCK & CO INC.
XX
XX Marsh DJ;
XX
XX WPI; 2001-565791/63.
DR N-PSDB; AAH47304.
XX
XX Fusion proteins comprising melanin concentrating hormone receptor
PT peptides and fluorescent proteins, useful for identifying appetite
PT stimulants.
XX
XX Claim 2; Page 14; 71pp; English.
XX
CC The invention provides melanin concentrating hormone (MCH) receptor
CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
CC MCHR polypeptide regions from different species. The MCHR fusion protein
CC comprise MCHR polypeptide region and a fluorescent polypeptide region
CC joined directly, or via a linker, to the carboxy side of the MCHR
CC polypeptide region. The MCHR fusion proteins can be expressed by standard
CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
CC regulation of MCH activity stimulates food intake. The present sequence
CC represents a A. victoria green fluorescent protein (GFP) and a linker
CC sequence
XX
XX Sequence 239 AA;
SQ

Query Match 98.7%; Score 1257; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 11
AAB31171
ID AAB31171 standard; protein; 239 AA.
XX
XX AAB31171;
AC
XX
XX 02-APR-2001 (first entry)
DT
XX
DE Amino acid sequence of a green fluorescent protein (GFP).
XX
KW Growth rate; death rate; reporter gene; luminescent protein;
KW Fluorescent product; luciferase; green fluorescent protein; GFP.
XX
XX Aequorea victoria.
OS
XX WO200075367-A1.
PN
XX
PD 14-DEC-2000.
XX
XX 07-JUN-2000; 2000WO-FI000507.
PF

XX
PR 07-JUN-1999; 99FI-00001296.
XX
XX (LILI/) LILIUS E.
PA (VIRT/) VIRTA M.
XX
XX Lilius E, Virta M;
PI
XX WPI; 2001-061737/07.
DR N-PSDB; AAC86954.
XX
XX Assessing growth and death rates of a micro-organism in a desired
PT environment, by introducing 2 reporter genes encoding luminescent and
PT fluorescent products and detecting luminescent fluorescence.
XX
XX Disclosure; Page 27; 32pp; English.
XX
CC The specification describes a method for assessing the growth rate and
CC death rate of a micro-organism within a predetermined time period in a
CC desired environment. The method comprises introducing at least two
CC reporter genes encoding luminescent and/or fluorescent products into the
CC micro-organisms, incubating the micro-organism within the desired
CC environment, and detecting luminescence and/or fluorescence after a
CC predetermined time period. Use of two different markers within a micro-
CC organism enables the differentiation between growth and death rates. The
CC method is used to assess the growth rate and death rate of a micro-
CC organism within a predetermined time period in a desired environment. The
CC present sequence represents a green fluorescent protein (GFP), and is
CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
CC and is used in the method of the invention
XX
XX Sequence 239 AA;
SQ

Query Match 98.7%; Score 1257; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 12
AAG66198
ID AAG66198 standard; protein; 239 AA.
XX
XX AAG66198;
AC
XX
XX 17-JUN-2002 (first entry)
DT
XX
DE A. victoria green fluorescent protein (EGFP).
XX
KW Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
KW Green fluorescent protein; EGFP.
XX
XX Aequorea victoria.
OS
XX JP2002045189-A.
PN
XX
XX 12-FEB-2002.
PD
XX

```
PF 04-AUG--2000; 2000JP-00237165.
XX
PR 04-AUG--2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
PA
XX WPI; 2002-299190/34.
DR N-PSDB; ABL40628.
XX
XX A gene encoding cyan-green fluorescent protein.
PT
XX
XX Example; Page 14; 20pp; Japanese.
PS
XX The invention relates to a gene encoding proteins having cyan-green
CC fluorescence characteristic and having a function of showing stable
CC fluorescence characteristic in acid region. A method for the preparation
CC of a cyan-green fluorescent protein is provided which involves a
CC transformant transformed by a recombinant vector comprising the gene,
CC where the transformant is cultured and the protein is collected from the
CC culture. The present sequence represents the A. victoria green
CC fluorescent protein (EGFP)
XX
SQ Sequence 239 AA;
Query Match 98.7%; Score 1257; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKFICTTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKFICTTGKLPVWPWT 60
QY 51 LVTLSSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
Db 51 LVTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 131 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
RESULT 13
ABG94444
ID ABG94444 standard; protein; 239 AA.
XX
XX ABG94444;
XX
XX 27-NOV-2002 (first entry)
XX
DE Protease biosensor signal sequence #6.
XX
KW Detection; classification; identification; toxin detection; protease;
KW ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
KW toxic threat agent.
XX
XX Synthetic.
XX
XX US6416959-B1.
XX
XX 09-JUL-2002.
XX
XX 25-FEB-2000; 2000US-00513783.
XX
XX 27-FEB-1997; 97US-00810983.
XX
XX 27-FEB-1998; 98US-00031271.
XX
XX 26-FEB-1999; 99US-0122152P.
XX
XX 08-MAR-1999; 99US-0123399P.
XX
XX 12-JUL-1999; 99US-00352171.
XX
XX 31-AUG-1999; 99US-0151797P.
XX
PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
PR 01-DEC-1999; 99US-0168408P.
XX
XX (GIUL/) GIULIANO K.
PA (KAPU/) KAPUR R.
XX
XX Giuliano K, Kapur R;
PI
XX WPI; 2002-634730/68.
DR N-PSDB; ABS71491.
XX
XX Automated cell-based toxin detection, classification, and/or
PT identification by treating cells involves use of three classes of
PT luminescent reporter molecules such as detectors, classifiers or
PT identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
PS
XX The invention describes methods of automated detection, classification
CC and identification comprising treating cells containing luminescent
CC reporter molecules (i) in array of locations with a test substance, where
CC (i) are detectors, classifiers or identifiers, imaging cells in each
CC location to obtain luminescent signals and converting optical information
CC into digital data to interpret presence of toxins in the test substance.
CC The method are useful for detection of toxins chosen from proteases, ADP-
CC ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
CC Three classes of cell-based luminescent reporter molecules such as
CC detectors, classifiers and identifiers are described and serve as
CC reporters of toxic threat agents. The first two levels of
CC characterisation ensure a rapid readout of toxin class without
CC sacrificing the ability to detect many new mutant toxins or dissect
CC several complex mixtures of known toxins. This is the amino acid sequence
CC of a protease biosensor related signal sequence used in the cell-based
CC screening system
XX
SQ Sequence 239 AA;
Query Match 98.7%; Score 1257; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKFICTTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKFICTTGKLPVWPWT 60
QY 51 LVTLSSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
Db 51 LVTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 131 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AAE14599;
XX
XX 31-MAY-2002 (first entry)
XX
DE Aequorea victoria enhanced green fluorescent protein.
XX
XX Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; muten.
XX
XX Aequorea victoria.
OS Synthetic.
```

XX Key Location/Qualifiers
 FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"
 FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"
 FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"
 XX EP1178109-A1.
 XX 06-FEB-2002.
 XX 03-AUG-2001; 2001EP-00306650.
 PR 04-AUG-2000; 2000JP-00237166.
 XX (RIKE) RIKEN KK.
 XX Miyawaki A, Sawano A;
 PI WPI; 2002-208112/27.
 DR N-PSDB; AAD27910.
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.
 PS Example 1; Page 13-14; 31pp; English.
 XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria
 XX Sequence 239 AA;
 Query Match 98.7%; Score 1257; DB 5; Length 239;
 Best Local Similarity 98.7%; Pred. No. 2.7e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 DB 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
 DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGDIDFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 DB 121 VNRIELKGDIDFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 RESULT 15
 AAE34958
 ID AAE34958 standard; protein; 239 AA.
 XX AAE34958;
 AC AAE34958;
 XX 28-MAY-2003 (first entry)
 XX Aequorea victoria enhanced green fluorescent protein (EGFP).
 XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;
 KW kinase; enhanced green fluorescent protein; EGFP.
 XX Aequorea victoria.
 OS WO200295058-A2.
 PN 28-NOV-2002.
 XX 24-MAY-2002; 2002WO-US016955.
 XX 24-MAY-2001; 2001US-00865291.
 PR (REGC) UNIV CALIFORNIA.
 PA Tsien RY, Ting AY, Zhang J;
 XX WPI: 2003-148474/14.
 DR N-PSDB; AAD53428.
 XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.
 PT Disclosure; Col 56-57; 38pp; English.
 PS The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or in operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention
 XX Sequence 239 AA;
 Query Match 98.7%; Score 1257; DB 6; Length 239;
 Best Local Similarity 98.7%; Pred. No. 2.7e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 DB 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
 DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGDIDFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 DB 121 VNRIELKGDIDFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

Sun Jun 27 18:27:37 2004

Db 131 DHYQNTPIGCGFVLLPDNNHLSOSALS KDPNEKRDHNVLLFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:10
Job time : 48.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-64L
Perfect score: 1273
Sequence: 1 MWSKGEELFTGVVPIILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1257	98.7	239	3	US-09-172-063-3
2	1257	98.7	239	4	US-09-513-783A-46
3	1257	98.7	239	4	US-09-316-919-4
4	1257	98.7	239	4	US-09-602-641-3
5	1257	98.7	239	4	US-09-920-922-2
6	1257	98.7	281	3	US-09-062-102-1
7	1257	98.7	281	4	US-09-364-946-1
8	1257	98.7	294	4	US-09-513-783A-2
9	1257	98.7	323	3	US-09-172-063-21
10	1257	98.7	323	4	US-09-602-641-21
11	1257	98.7	364	3	US-09-085-305-6
12	1257	98.7	379	4	US-09-417-197-129
13	1257	98.7	434	4	US-09-800-170-48
14	1257	98.7	442	4	US-09-417-197-127
15	1257	98.7	459	4	US-09-513-783A-170
16	1257	98.7	544	4	US-09-417-197-113
17	1257	98.7	544	4	US-09-417-197-115
18	1257	98.7	604	4	US-09-417-197-59
19	1257	98.7	605	4	US-09-417-197-41
20	1257	98.7	606	4	US-09-417-197-65
21	1257	98.7	607	4	US-09-417-197-47
22	1257	98.7	630	4	US-09-417-197-63
23	1257	98.7	631	4	US-09-417-197-39
24	1257	98.7	633	4	US-09-417-197-45
25	1257	98.7	635	4	US-09-417-197-125
26	1257	98.7	642	2	US-08-818-253-2
27	1257	98.7	642	2	US-08-818-253-6

28	1257	98.7	642	3	US-08-818-252-2	Sequence 2, Appli
29	1257	98.7	642	3	US-08-818-252-6	Sequence 6, Appli
30	1257	98.7	652	3	US-08-818-253-4	Sequence 4, Appli
31	1257	98.7	652	3	US-08-818-252-4	Sequence 4, Appli
32	1257	98.7	718	4	US-09-417-197-75	Sequence 75, Appli
33	1257	98.7	719	4	US-09-417-197-51	Sequence 51, Appli
34	1257	98.7	726	4	US-09-417-197-71	Sequence 71, Appli
35	1257	98.7	727	4	US-09-417-197-139	Sequence 139, App
36	1257	98.7	783	4	US-09-513-783A-176	Sequence 176, App
37	1257	98.7	797	4	US-09-417-197-141	Sequence 141, App
38	1257	98.7	797	4	US-09-417-197-143	Sequence 143, App
39	1257	98.7	798	4	US-09-417-197-77	Sequence 77, Appli
40	1257	98.7	805	4	US-09-513-783A-178	Sequence 178, App
41	1257	98.7	806	4	US-09-417-197-53	Sequence 53, Appli
42	1257	98.7	836	4	US-09-417-197-61	Sequence 61, Appli
43	1257	98.7	842	4	US-09-417-197-43	Sequence 43, Appli
44	1257	98.7	843	4	US-09-417-197-117	Sequence 117, App
45	1257	98.7	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1
US-09-172-063-3
; Sequence 3, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Liopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-172-063-3

Query Match	98.7%;	Score 1257;	DB 3;	Length 239;
Best Local Similarity	98.7%;	Pred. No. 9.3e-127;		
Matches 236;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	MYSKGEELFTGVVPIILVELDGDVNGHKFVS	SGEGDATYVKLTAKFICTTGKLPVPWPT	60
Db	1	MYSKGEELFTGVVPIILVELDGDVNGHKFVS	SGEGDATYVKLTAKFICTTGKLPVPWPT	60
QY	61	LVTLLSYGVQCFSRYPDHHKQHDFFKSA	MEGVVQBERTIFFKDDGNYKTRAEVKFEGDTL	120
Db	61	LVTLLTYGVQCFSRYPDHHKQHDFFKSA	MEGVVQBERTIFFKDDGNYKTRAEVKFEGDTL	120
QY	121	VNRIELKGDIFDKEDGNILGHKLEYNYSNHNVI	YIMADKQNGIKVNFIRHNIEDGSVOLA	180
Db	121	VNRIELKGDIFDKEDGNILGHKLEYNYSNHNVI	YIMADKQNGIKVNFIRHNIEDGSVOLA	180
QY	181	DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDP	NEKRDHVVLLGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDP	NEKRDHVVLLGFVTAAGITLGMDELYK	239

QY 121 VNRILKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRILKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239

RESULT 4
 US-09-602-641-3
 ; Sequence 3, Application US/09602641
 ; Patent No. 6608189
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Miyawaki, Atsushi
 ; APPLICANT: Llopis, Juan
 ; APPLICANT: Wachter, Rebekka M.
 ; APPLICANT: Remington, S. James
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
 ; MEASURING THE PH OF A BIOLOGICAL SAMPLE
 ; FILE REFERENCE: 07257/071001
 ; CURRENT APPLICATION NUMBER: US/09/602,641
 ; PRIOR FILING DATE: 2000-06-22
 ; PRIOR APPLICATION NUMBER: 09/172,063
 ; PRIOR FILING DATE: 1998-10-13
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 239
 ; TYPE: PRT
 ; ORGANISM: Aequorea victoria
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (0)...(0)
 ; OTHER INFORMATION: EGFP
 US-09-602-641-3

Query Match 98.7%; Score 1257; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 9.3e-127;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRILKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRILKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239

RESULT 5
 US-09-920-922-2
 ; Sequence 2, Application US/09920922
 ; Patent No. 6673610
 ; GENERAL INFORMATION:
 ; APPLICANT: Miyawaki, Atsushi
 ; APPLICANT: Sawano, Asako
 ; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
 ; FILE REFERENCE: 11283-012001
 ; CURRENT APPLICATION NUMBER: US/09/920,922
 ; CURRENT FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: JP 2000-237166
 ; PRIOR FILING DATE: 2000-08-04
 ; NUMBER OF SEQ ID NOS: 9

RESULT 2
 US-09-513-783A-46
 ; Sequence 46, Application US/09513783A
 ; Patent No. 6416959
 ; GENERAL INFORMATION:
 ; APPLICANT: Giuliano, Kenneth A.
 ; APPLICANT: Kapur, Ravi
 ; TITLE OF INVENTION: A System for Cell Based Screening
 ; FILE REFERENCE: 97-022-L1
 ; CURRENT APPLICATION NUMBER: US/09/513,783A
 ; CURRENT FILING DATE: 2000-02-25
 ; NUMBER OF SEQ ID NOS: 180
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 46
 ; LENGTH: 239
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: EGFP
 US-09-513-783A-46

Query Match 98.7%; Score 1257; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 9.3e-127;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRILKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRILKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239

RESULT 3
 US-09-316-919-4
 ; Sequence 4, Application US/09316919
 ; Patent No. 6469154
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Baird, Geoffrey
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
 ; FILE REFERENCE: 07257/073001
 ; CURRENT APPLICATION NUMBER: US/09/316,919
 ; CURRENT FILING DATE: 1999-05-21
 ; NUMBER OF SEQ ID NOS: 63
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 239
 ; TYPE: PRT
 ; ORGANISM: Aequorea victoria
 US-09-316-919-4

Query Match 98.7%; Score 1257; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 9.3e-127;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.7%; Score 1257; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 9.3e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Qy 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          98.7%; Score 1257; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.2e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Qy 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CIP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          98.7%; Score 1257; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.2e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Qy 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSQKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match          98.7%; Score 1257; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60

```

Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60
QY 51 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 51 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9

US-09-172-053-21
; Sequence 21, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-172-053-21

Query Match 98.7%; Score 1257; DB 3; Length 323;
Best Local Similarity 98.7%; Pred. No. 1.5e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60
Db 35 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 144
QY 51 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 204
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRIELKGIDFKEDGNILGHKLEYNHNHVMYIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 255 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21
; Sequence 21, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-602-641-21

Query Match 98.7%; Score 1257; DB 4; Length 323;
Best Local Similarity 98.7%; Pred. No. 1.5e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60
Db 85 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 144
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 204
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRIELKGIDFKEDGNILGHKLEYNHNHVMYIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6
; Sequence 6, Application US/09085305
; Patent No. 6191269
; GENERAL INFORMATION:
; APPLICANT: Pollock, Allan
; APPLICANT: Lovett, David H.
; APPLICANT: Turck, Johanna
; TITLE OF INVENTION: Selective Induction of Apoptosis in
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,305
; FILING DATE: 29-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Francis, Carol L
; REGISTRATION NUMBER: 36,513
; REFERENCE/DOCKET NUMBER: 6510/102US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-085-305-6

Query Match          98.7%; Score 1257; DB 3; Length 364;
Best Local Similarity 98.7%; Pred. No. 1.8e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 126 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 185
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 186 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 245
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 246 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 305
QY 181 DHYQQNTPIGDGVVLPDNNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 306 DHYQQNTPIGDGVVLPDNNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 364

RESULT 12
US-09-417-197-129
; Sequence 129, Application US/09417197
; Patent No. 6518021
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP, et al.
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
; FILE OF INVENTION: On A Cellular Response
; FILE REFERENCE: 3759-0110P
; CURRENT APPLICATION NUMBER: US/09/417,197
; CURRENT FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 129
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: actin-binding-domain-EGFP fusion
US-09-417-197-129

Query Match          98.7%; Score 1257; DB 4; Length 379;
Best Local Similarity 98.7%; Pred. No. 1.9e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 141 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 200
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 201 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 260
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 261 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 320
```

```
QY 181 DHYQQNTPIGDGVVLPDNNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 321 DHYQQNTPIGDGVVLPDNNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 379

RESULT 13
US-09-800-170-48
; Sequence 48, Application US/09800170
; Patent No. 6481667
; GENERAL INFORMATION:
; APPLICANT: Kinsella, Todd
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/800,170
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Synechocystis PCC6803
US-09-800-170-48

Query Match          98.7%; Score 1257; DB 4; Length 434;
Best Local Similarity 98.7%; Pred. No. 2.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 196 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 255
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 256 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 315
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 316 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 375
QY 181 DHYQQNTPIGDGVVLPDNNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 376 DHYQQNTPIGDGVVLPDNNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 434
```

```
RESULT 14
US-09-417-197-127
; Sequence 127, Application US/09417197
; Patent No. 6518021
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP, et al.
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
; FILE OF INVENTION: On A Cellular Response
; FILE REFERENCE: 3759-0110P
; CURRENT APPLICATION NUMBER: US/09/417,197
; CURRENT FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 127
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-RhoA fusion
US-09-417-197-127

Query Match          98.7%; Score 1257; DB 4; Length 442;
Best Local Similarity 98.7%; Pred. No. 2.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
```

|||||
Db 1 MVSKEELFTGVVPIVLVDGVNGHKFSVSGEGDATYGLTLKFCITCTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 641659
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPP-HSP27
US-09-513-783A-170

Query Match 98.7%; Score 1257; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 2.5e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVNGHKFSVSGEGDATYGLTLKFCITCTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPIVLVDGVNGHKFSVSGEGDATYGLTLKFCITCTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:02
Job time : 13.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds
(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-64L
Perfect score: 1273
Sequence: 1 MVSKEELFTGVVPIVLD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pap.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pap.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pap.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pap.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1268	99.6	239	9	US-09-887-784-4
2	1268	99.6	239	12	US-10-296-953-4
3	1268	99.6	239	14	US-10-270-223-6
4	1268	99.6	893	14	US-10-257-909A-30
5	1268	99.6	1132	14	US-10-257-909A-32
6	1260	99.0	239	9	US-09-887-784-2
7	1260	99.0	239	12	US-10-296-953-2
8	1257	98.7	239	9	US-09-920-922-2
9	1257	98.7	239	9	US-09-999-745-4
10	1257	98.7	239	10	US-09-866-538-4
11	1257	98.7	239	10	US-09-797-496B-2
12	1257	98.7	239	10	US-09-794-308-4
13	1257	98.7	239	10	US-09-865-291-4
14	1257	98.7	239	12	US-10-457-982-3
15	1257	98.7	239	14	US-10-121-258-13

16	1257	98.7	239	14	US-10-221-461-7	Sequence 7, Appli
17	1257	98.7	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1257	98.7	239	14	US-10-177-390-2	Sequence 2, Appli
19	1257	98.7	239	14	US-10-338-411-3	Sequence 3, Appli
20	1257	98.7	239	15	US-10-370-570-4	Sequence 4, Appli
21	1257	98.7	239	15	US-10-389-640-3	Sequence 3, Appli
22	1257	98.7	259	14	US-10-314-861-11	Sequence 11, Appli
23	1257	98.7	281	12	US-09-931-232-1	Sequence 1, Appli
24	1257	98.7	288	14	US-10-314-861-37	Sequence 37, Appli
25	1257	98.7	293	14	US-10-314-861-35	Sequence 35, Appli
26	1257	98.7	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1257	98.7	295	14	US-10-314-861-39	Sequence 39, Appli
28	1257	98.7	299	14	US-10-314-861-33	Sequence 33, Appli
29	1257	98.7	305	14	US-10-314-861-31	Sequence 31, Appli
30	1257	98.7	308	14	US-10-033-717-35	Sequence 35, Appli
31	1257	98.7	311	14	US-10-314-861-29	Sequence 29, Appli
32	1257	98.7	320	14	US-10-338-411-11	Sequence 11, Appli
33	1257	98.7	320	15	US-10-389-640-11	Sequence 11, Appli
34	1257	98.7	323	12	US-10-457-982-21	Sequence 21, Appli
35	1257	98.7	323	14	US-10-338-411-7	Sequence 7, Appli
36	1257	98.7	323	14	US-10-338-411-13	Sequence 13, Appli
37	1257	98.7	323	15	US-10-389-640-7	Sequence 7, Appli
38	1257	98.7	323	15	US-10-389-640-13	Sequence 13, Appli
39	1257	98.7	324	14	US-10-314-861-16	Sequence 16, Appli
40	1257	98.7	345	14	US-10-338-411-5	Sequence 5, Appli
41	1257	98.7	345	15	US-10-389-640-5	Sequence 5, Appli
42	1257	98.7	346	14	US-10-338-411-9	Sequence 9, Appli
43	1257	98.7	346	15	US-10-389-640-9	Sequence 9, Appli
44	1257	98.7	359	14	US-10-033-717-33	Sequence 33, Appli
45	1257	98.7	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequoria Victoria
US-09-887-784-4

Query Match 99.6%; Score 1268; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 4e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKGEELFTGVVPIVLDGVDVNGHKFSVSGEGDATYGKLT	FKFICTTGKLPVPWPT 60
Db	1	MVSKGEELFTGVVPIVLDGVDVNGHKFSVSGEGDATYGKLT	FKFICTTGKLPVPWPT 60
Qy	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSA	MPGVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSA	MPGVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Qy	121	VNRIELKIDFDKEDGNILGHKLEYNHN	YIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db	121	VNRIELKIDFDKEDGNILGHKLEYNHN	YIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Qy	181	DHYQNTPTIGDGPVLLPDNHYLSTQSA	LKSDPNKRDHNVLLGFVTAAGITLGMDELYK 239
Db	181	DHYQNTPTIGDGPVLLPDNHYLSTQSA	LKSDPNKRDHNVLLGFVTAAGITLGMDELYK 239

```
RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match          99.6%; Score 1268; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 4e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 3
US-10-270-223-6
; Sequence 6, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: BioImage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM INTERACTIONS BY FLUORESCENCE REDISTRIBUTION.
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match          99.6%; Score 1268; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 7.4e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.6%; Score 1268; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 2.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 655 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 714
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.6%; Score 1268; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 2.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 655 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 714
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
```



```

; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.6%; Score 1268; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 3.8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 953
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          99.0%; Score 1260; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          98.7%; Score 1257; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
```

```

; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.0%; Score 1260; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.7%; Score 1257; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSFGSEGDATYVKLTLPKICTTGKLPVPWPT 60
```


APPLICANT: TSIEN, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match 98.7%; Score 1257; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGVPLLPDNHNLSTQSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPLLPDNHNLSTQSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
Sequence 4, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEN, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REGEN1550
CURRENT APPLICATION NUMBER: US/09/865,291
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match 98.7%; Score 1257; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US20030212265A1
GENERAL INFORMATION:
APPLICANT: Tsiien, Roger Y.
APPLICANT: Miyawaki, Atsushi
APPLICANT: Liopis, Juan
APPLICANT: Wachter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR APPLICATION NUMBER: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 98.7%; Score 1257; DB 12; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGVPLLPDNHNLSTQSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPLLPDNHNLSTQSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 15
US-10-121-258-13
Sequence 13, Application US/10121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Campbell, Robert
APPLICANT: Tsiien, Roger
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP1
CURRENT APPLICATION NUMBER: US/10/121,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24

QY 181 DHYQONTPIGDGVPLLPDNHNLSTQSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPLLPDNHNLSTQSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US20030212265A1
GENERAL INFORMATION:
APPLICANT: Tsiien, Roger Y.
APPLICANT: Miyawaki, Atsushi
APPLICANT: Liopis, Juan
APPLICANT: Wachter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR APPLICATION NUMBER: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 98.7%; Score 1257; DB 12; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGVPLLPDNHNLSTQSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPLLPDNHNLSTQSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.7%; Score 1257; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1  MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYVKLTLPFICTTGKLPVPWPT 60
Db      1  MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYVKLTLPFICTTGKLPVPWPT 60

QY     51  LVTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db     51  LVTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY     131  DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLLGFVTAAGITLGMDELYK 239
Db     131  DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLLGFVTAAGITLGMDELYK 239
```

Search completed: June 21, 2004, 16:09:27
Job time : 34.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds
(without alignments)
2224.817 Million cell updates/sec

Title: US-09-887-784-64L

Perfect score: 1273

Sequence: 1 MVSKEBELTGVVPIVLVELD.....VLLGFVTAAGITLGMDELK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1235	97.0	238	1 JQ1514	green-fluorescent
2	1105	8.2	785	2 H72228	hypothetical prote
3	96.5	7.6	861	2 H64102	leucine-trna ligas
4	94.5	7.4	887	2 E82590	leucyl-trna synth
5	92.5	7.3	860	2 AC0582	leucyl-trna synth
6	87.5	6.9	655	2 D83917	DNA topoisomerase
7	87.5	6.9	2573	2 D71614	hypothetical prote
8	86.5	6.8	370	2 E70390	iron-sulfur cofact
9	86.5	6.8	788	1 JDVLHH	DNA-directed DNA p
10	86.5	6.8	787	2 JC4078	protective surface
11	86.5	6.8	808	2 F64102	protective surface
12	86.5	6.8	941	2 S29043	cellulase (EC 3.2.
13	86	6.8	357	2 G81355	trna (uracil-5)-m
14	85.5	6.7	860	1 SYECL	leucine-trna ligas
15	85.5	6.7	860	2 H90713	leucine-trna synth
16	85.5	6.7	860	2 D85564	leucine trna synth
17	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
18	85	6.7	281	2 AD2052	hypothetical prote
19	85	6.7	411	2 T06936	photosystem II chl
20	84.5	6.6	613	2 A99552	oligoendopeptidase
21	84.5	6.6	836	1 JDHVL	DNA-directed DNA p
22	84	6.6	439	2 JH0414	synaptogamin o-p65
23	84	6.6	529	2 B86815	hypothetical prote
24	84	6.6	578	1 I40794	dihydrolipoamide d
25	84	6.6	622	2 T06586	DNA-binding protei
26	84	6.6	752	1 KXRTC1	proprotein convert
27	83.5	6.6	435	2 C97354	hypothetical prote
28	83.5	6.6	860	2 AF0318	leucine-trna ligas
29	83.5	6.6	1134	2 A60234	Iga Fc receptor pr

30	83.5	6.6	1164	1 FCSOAG	Iga Fc receptor pr
31	83	6.5	285	2 C97279	thioredoxin reduct
32	83	6.5	836	2 T42323	hypothetical prote
33	82	6.4	336	2 C64468	hypothetical prote
34	82	6.4	353	2 E84941	imidazoleglycerol-
35	82	6.4	355	2 H97144	ferriochrome-binding
36	82	6.4	471	2 T27856	hypothetical prote
37	82	6.4	774	2 T39539	alpha-amylase homo
38	82	6.4	865	2 T11852	lipoxigenase (EC 1
39	81.5	6.4	263	2 S53488	water-stress-induc
40	81.5	6.4	2222	1 A36028	DNA-directed DNA p
41	81	6.4	272	2 B71094	hypothetical prote
42	81	6.4	310	2 S68225	synergolymenotropi
43	81	6.4	459	2 S42647	photosystem II chl
44	81	6.4	459	2 AD2342	photosystem II CP4
45	81	6.4	559	2 A45620	cytoவில்lin homolog

ALIGNMENTS

RESULT 1

JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:1347277
A:Accession: JQ0692
A:Molecule type: DNA
A:Residues: 1-107, 'S', 109-238 <PRA1>
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>
A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384
R:Watkins, J.N.; Campbell, A.K.
Submitted to the EMBL data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13, 'V', 15-24, 'Q', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009
A:Experimental source: clone gfp1
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 209-238 <PRA5>
A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011
A:Experimental source: clone gfp2
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65692; PDB:1GFL
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-91;
A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A59953; MUID:98294543; PMID:9631087

A;Contents: annotation; X-ray crystallography, 1.9 angstroms
C;Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
C;Genetics:

A;Gene: GFP
A;Introns: 59/3; 167/3
C;Superfamily: green-fluorescent protein
C;Keywords: Chromoprotein; luminescence
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.0%; Score 1235; DB 1; Length 238;
Best Local Similarity 96.6%; Pred. No. 1.6e-96;
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```
QY 2 VSKGEELFTGVVPIVLVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGVVPIVLVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 52 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 121
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 51 VITFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIIFKDDGNYKTRAEVKFEQDTLV 120
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFIRHNIEDGSVQLAD 181
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGIDFKEDGNILGHKMEYNYNHNVIMADKQNGIKVNFIRHNIEDGSVQLAD 180
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 192 HYQONTPIGDGPVLLPDNHYLSQTQSALSKDPNKRDMHVLGFTVTAAGITGLMDELYK 239
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 HYQONTPIGDGPVLLPDNHYLSQTQSALSKDPNKRDMHVLLEFVTAAGITGHMDELYK 238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 2

H72228
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: H72228
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: H72228

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-785 <ARN>
A;Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219
A;Experimental source: strain MSB8
C;Genetics:

A;Gene: TM1524

Query Match 8.2%; Score 105; DB 2; Length 785;
Best Local Similarity 19.7%; Pred. No. 0.84;
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

```
QY 3 SKGEELFTGVVPIVLVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTLV 62
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 15 NEGRFSEGTVPGVQAD-----LVRKGLLPHPYVGM- 46
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 53 TLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLVN 122
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 47 -----NEDLFKEIDREWIYVERFEKDEKGERVDLVFEGVDTL 88
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 123 RIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFIRHNIEDGSVQLADH 182
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 39 DVYLVNGVYL--GSTEDMFIEYRFDVTNLV----KEKNHLKVYIK-----SPIRVPKT 134
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 193 YQONTPIGDGPVLLPDNHYLSQTQSALSKDPNKRDMHVLGFTVTAAGITGLMD 235
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 135 LEQNYGVLGSP-----EDP-----INGYIRKAQYSYGD 163
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 3

H64102

leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)
N;Alternate names: leucyl-tRNA synthetase
C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002

C;Accession: H64102

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C.
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: H64102

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-861 <TIGR>

A;Cross-references: GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC22581.1; PID:g1573943; T3

C;Genetics:

C;Superfamily: leucine-tRNA ligase

C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.6%; Score 96.5; DB 2; Length 861;

Best Local Similarity 24.6%; Pred. No. 4.9;

Matches 47; Conservative 26; Mismatches 59; Indels 59; Gaps 11;

```
QY 50 TTGKLPVPWPTLVTLSSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYK 103
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 314 TGDKLPI--WVANFVLMHYGTGAVMAVPAH--DQRDF-----EFAQKYSLPKQVIAPLA 364
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 104 DGNKTRAEVKSEGDTLVNRIELKGDIDFKEDGNILGHKLEYNYNHNVIMADK-QKNGI 162
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 365 DEEIDLTQAFVEHGHKLVNSDFDGKNF--DGAENG-----IADLKEKLG 408
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 163 ---KVNFKIRH-----NIEDGSVQLADHYQQNTPIGDPVLLPDNHYL- 202
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 409 GKQVNYRLRDGWGVSQRVYWGAPIPMLTLENGDVVPA-----PMEDLPILIPEDVWMD 461
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 4

E82590

leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)

C;Species: Xylella fastidiosa

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C;Accession: E82590

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: E82590

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-887 <SIM>

A;Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN0012

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
Briones, M.R.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H.
as-Neto, E.; Docena, C.; El-Dorry, H.; Facinani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z M.; Reference number: A59328
A;Content: annotation
C;Genetics:
A;Gene: XP2176
C;Superfamily: leucine-tRNA ligase

Query Match 7.4%; Score 94.5; DB 2; Length 887;
Best Local Similarity 22.7%; Pred. No. 7.5;
Matches 45; Conservative 29; Mismatches 67; Indels 57; Gaps 10;

QY 50 TTGKLPVNPPTLVTLSSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNY-- 107
Db TNEQLPV-WVANFVLMAYGTGAVMAVPGHQRQDEF--ANKYGLPIRQVIALKEPKNQDB 385

QY 108 -----KTRAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYNNSHNVI 153
Db STWEPDVRDWDYADKTR---EFE---LINSAPFDGLDQDAFEVLAERFE----- 429

QY 154 MADKQKNG-IKNFKIRHNIEDGSVOLADHYQOQNTPI-----GDGPVLLPDN 199
Db ---RQGRQRRVNYRLR-----DWGVSQRQYWGCPPIVIVCTCGAVPVEDQLPVILPEN 482

QY 200 -HYLSQTOSALSKDPEK 216
Db VAPSGTGTGSPIKTPDPEWK 500

RESULT 5
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
R;Accession: AC0582
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
C;Accession: AC0582
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-860 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:gl6501899; GSPDB:GN00176
C;Genetics:
A;Gene: STY0699
C;Superfamily: leucine-tRNA ligase

Query Match 7.3%; Score 92.5; DB 2; Length 860;
Best Local Similarity 23.9%; Pred. No. 11;
Matches 43; Conservative 21; Mismatches 77; Indels 39; Gaps 7;

QY 50 TTGKLPVNPPTLVTLSSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKT 109
Db TGEELPV-WVANFVLMAYGTGAVMAVPGH-QDQD-YEFASKYGLTIKPVILLADGSEPD 370

QY 110 RAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIR 169
Db SEQALTEKGVLFNSGFDFGLAFEAFAFNAIADKL-----AEKVGGERKVNRYLR 418

QY 170 H-----NIEDGSVOLADHYQOQNTPIGDGPVLLPDNHYL-STQSALS KDP 212
Db DWGVSQRQYWGAPIPVMTLEDGT-----LPTPEDQLPVILPDVDMVDGITSPIKADP 471

RESULT 6
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: D83917
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
C;Accession: D83917
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <STO>
A;Cross-references: GB:AP001514; GB:BA000004; NID:gl0174613; PIDN:BA05859.1; GSPDB:GN001
C;Experimental source: strain C-125
C;Gene: BH2140
C;Superfamily: DNA topoisomerase (Atp-hydrolyzing) chain B

Query Match 6.9%; Score 87.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 20;
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK---FSVSGEGEGDAT---YGLTLTKFI-----CTTGKLPVPWP 59
Db 63 NVTIHKQDSVVRDEGRGMPGTGMHKLKCKPTPEVILTVLHAGGKFGGGYATSGLHGVA 122

QY 60 TLVTLSSYGVCFSRYPDHMKQHDFFKSAPEGYVOER-----TIFKDDG----- 105
Db 123 SVVNALSEWLIVEIKRDGHWYEQRPENGKGPSTTLEKKGKTRGTGTHFKPDPTVFSTT 182

QY 106 --NYKTRAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYNNSHNVIYIMADK----- 157
Db 183 NFNVTLSERLEAAFLKGLKIELVLDLDDTKEVPH-YEDGKAFVEYLNEKDLHPV 241

QY 158 -----QKNGIKVNFKIRHNIEDGSVOLADHYQOQNTPIGDGPVLLPDNHYLSTQSALS KDP 212
Db 242 VFNESNGIEIEFAPQFN--DGYENVLSFVNNVTKDGG-----GTHLGAKTAMTRAV 294

QY 213 NE 214
Db 295 NE 296

RESULT 7
D71614
hypothetical protein PF0460c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: D71614
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
C;Accession: D71614
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2573 <GAR>
A;Cross-references: GB:AE001396; GB:AE001362; NID:g3845188; PIDN:AACT71881.1; PID:g384519
C;Experimental source: clone 3D7
C;Genetics:
A;Gene: PF0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;
Best Local Similarity 26.2%; Pred. No. 1.2e+02;
Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VOERTIFFKDD--DGNKYKTRAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYN--YN 149
Db 126 LKKTILCKDIKSGNDPMDIEISLFKDDMVDDKELK--DFEKSLSLKIKNKEVYNYNKM 183

QY 150 NYTIMADKQKNGIKVNFKIRHNIEDGSVOLADHYQOQNTPIGDGPVLLPDNHYLSTQSALS 209
Db 184 NLHAIKENKKDEKKNKHNNDNNM---IYKNI---DKTHYILDNNVVHILNDIN 236

```
QY 210 KDPNEKRDHM 219
Db 237 TVLKRERDYM 246

RESULT 8
E70390
N;Contains: iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999
C;Accession: E70390
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oviatt, R.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70390
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-370 <AQF>
A;Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: nifs1
C;Superfamily: nitrogen fixation protein nifs
C;Keywords: phosphoprotein; pyridoxal phosphate; sulfotransferase
F;195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted
F;318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match 6.8%; Score 86.5; DB 2; Length 370;
Best Local Similarity 25.4%; Pred. No. 11;
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY 4 KBELFTGVV----PILVELD----GDVNGHKF-SVSGEG----EGDATYVGLTLKFICT 50
Db 164 KGVPLLTDAVQAIKPIELKNISYATFSCHKFHAIKSGSFLYISDEANVEPLIVGGQE 223

QY 51 TGKLP-----VPMPTLVLLSYGVQCFSPYDPHMKQ-HDFFKAMPEGYVQERTIFPKDD 104
Db 224 NGRSGTENNVGLISLAKALEIIVNFSRYQEQKKLRDLFLNLLLEA-LPDAQIVGKOA 282

QY 105 GNYKTRAEV---KFGDGLVRELKIDPKEDGNILGHKLYNNSHNHYIMADKQNG 161
Db 283 ERSFSISSVIMPFFGAEIVNKLSEKGIYCSGACLSGEYEFNKHMLKMGFSQERKALRM 342

QY 162 IKVNFKIRHNIED 174
Db 343 VRFSGLLNKEEE 355

RESULT 9
JDVLHH
DNA-directed DNA polymerase (EC 2.7.7.7) - heron hepatitis virus
C;Species: heron hepatitis virus, HHV9
A;Note: host Ardea cinerea (gray heron)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 11-Jun-1999
C;Accession: A30082
R;Sprengel, R.; Kaleta, E.F.; Will, H.
J. Virol. 62, 3832-3839, 1988
A;Title: Isolation and characterization of a hepatitis B virus endemic in herons.
A;Reference number: A93037; MUID:98333160; PMID:3418788
A;Accession: A30082
A;Molecule type: DNA
A;Residues: 1-788 <SPR>
A;Cross-references: GB:M22056; NID:g325452; PIDN:AAA45738.1; PID:g325454
C;Superfamily: hepatitis virus DNA-directed DNA polymerase
C;Keywords: DNA biosynthesis; nucleotidyltransferase

Query Match 6.8%; Score 86.5; DB 1; Length 788;
Best Local Similarity 19.7%; Pred. No. 30;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;
```

```
QY 58 WPTLVLLSYGVQCFSPYDPHMKQ-----DFFKAMPEGYVQERT-----IFFKDDGNKY 108
Db 139 WPKSISYLPVHSVGVPKYPFQQHESLVNDYLNKLFFAGILYKVSXHLVTFK--GPVF 196

QY 109 T-----RAEVKPEGDTLVNRIELKIDFKEDGNILGHKLYNNSHNHYIM 154
Db 197 TWEQKHLVPQOHGAYSSKINDRQESRRRIITATSSRKNDSSRI-----FGAHN---- 245

QY 155 ADKQKNGIKVNFKIRHNIEDSGVQLADHYQ-----QNTPIGDGPVLL--PDNHYL 202
Db 246 -----NGRKISY---HSTRDGHSLRSTSDPTSRGALAGDSTPIGPGSTAHPSTHHV 297

QY 203 STQ-----SALSKDPNEKR 216
Db 298 DRRRQKQGVVQLAISREPSETR 320

RESULT 10
JC4078
Protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C;Species: Haemophilus influenzae
A;Variety: type b
C;Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C;Accession: JC4078
R;Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A;Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl
A;Reference number: JC4078; MUID:95255676; PMID:7737523
A;Accession: JC4078
A;Molecule type: DNA
A;Residues: 1-797 <FLA>
A;Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448
A;Experimental source: type b
C;Superfamily: protective surface antigen D-15
C;Keywords: surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 6.8%; Score 86.5; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. No. 31;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY 65 LSYGVQCFSPYDPHMKQHDFF-----FKSAMPEGYVQE-----RTI 99
Db 427 IGVGTESGISYQASVKQDNFLGTGAASVSIAGTKNDYGTSVNLGYTEPYFTKDGVSJGNNV 486

QY 100 FPKDDGNYKTRAEVKFEGDTLVNRIELKIDFKEDGNI---LGH-----KLEYNNS 148
Db 487 PFENYDNSKSDTSSNYKRTTYGNSVTL-GFPVNNNSYVVGLOHTYKNKISNFALEYN--- 542

QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDSGVQLADHYQ-----NTPIGDGPVLL 196
Db 543 RNLIYQSMKFKNGIKTN-----DFDFSFGVNNLSLNRGYFTPKGVKASLG-GRVTI 593

QY 197 P--DNHYLSTQSAISKDPNEKRDHMLVLLGLFVTAAGITLG 233
Db 594 PGSDNKKYKLSADVOGFYPLDRDLHMLVWSAKASAGYANG 632

RESULT 11
F64102
Protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C;Accession: F64102
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: F64102
```


A;Accession: A30290
A;Molecule type: DNA
A;Residues: 1-66, 'H', 68-195, 'N', 197-261, 'R', 263-860 <HAE>
A;Cross-references: EMBL:X06331; NID:G41915; PIDN:CAA29642.1; PID:G41916
A;Note: part of this sequence, including the amino end of the mature protein, was confirmed
C;Genetics:
A;Gene: leuS
A;Map position: 15 min
C;Function:
A;Description: EC 6.1.1.4 [validated, MUID:88096562]
A;Pathway: protein biosynthesis
C;Superfamily: leucine-tRNA ligase
C;Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
F;49-52/Region: ATP binding #status predicted

Query Match 6.7%; Score 85.5; DB 1; Length 860;
Best Local Similarity 22.8%; Pred. No. 41;
Matches 41; Conservative 24; Mismatches 76; Indels 39; Gaps 8;

QY 50 TTGKLPVPWPTLVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109
DB 314 TGEIIPV-WAANFVLMVEYGTGAVNVPFGH-DQRD-YEFASKYGLNFKPVILAADGSEPD 370
QY 110 RAEVKFEGDGLVNRIELKGDIDFKEDGNILGHLEYNYNHNVYIMADKQKNGIKVNFKIR 169
DB 371 SQQALTEKGVLFNSGEFGLD-----HEAFAIAADKLTAMGVGER---KVNRYLR 418
QY 170 H-----NIEDGSVOLADHYQQNTPTGDPVLLPDNHYL-STQSALSKDP 212
DB 419 DWGVSQRQYWCAPIPMTVLEDGTV-----MPTDDQLPVILPEDVVMGDGITSPIKADP 471

RESULT 15
leucine tRNA synthetase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: H90713
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H90713
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-860 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAE34103.1; PID:gl3360138; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0680
C;Superfamily: leucine-tRNA ligase

Query Match 6.7%; Score 85.5; DB 2; Length 860;
Best Local Similarity 22.8%; Pred. No. 41;
Matches 41; Conservative 24; Mismatches 76; Indels 39; Gaps 8;

QY 50 TTGKLPVPWPTLVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109
DB 314 TGEIIPV-WAANFVLMVEYGTGAVNVPFGH-DQRD-YEFASKYGLNFKPVILAADGSEPD 370
QY 110 RAEVKFEGDGLVNRIELKGDIDFKEDGNILGHLEYNYNHNVYIMADKQKNGIKVNFKIR 169
DB 371 SQQALTEKGVLFNSGEFGLD-----HEAFAIAADKLTAMGVGER---KVNRYLR 418
QY 170 H-----NIEDGSVOLADHYQQNTPTGDPVLLPDNHYL-STQSALSKDP 212
DB 419 DWGVSQRQYWCAPIPMTVLEDGTV-----MPTDDQLPVILPEDVVMGDGITSPIKADP 471

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.4444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-64L

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPLVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1241	97.5	238	1	GFP_AEQVI
2	96.5	7.6	861	1	SYL_HA5IN
3	94.5	7.4	879	1	SYL_XYLFA
4	92.5	7.3	860	1	SYL_SALT
5	92.5	7.3	860	1	SYL_SALT
6	91.5	7.2	879	1	SYL_XYLFT
7	89.5	7.0	859	1	SYL_SHEON
8	88.5	7.0	538	1	GRBE_RAT
9	86.5	6.8	533	1	CP51_CANGA
10	86.5	6.8	788	1	DPOL_HPBHE
11	86.5	6.8	795	1	D152_HAEIN
12	86.5	6.8	797	1	D151_HAEIN
13	86.5	6.8	941	1	GUN_BACS6
14	86	6.8	357	1	TRMF_CAMJE
15	85.5	6.7	793	1	D153_HAEIN
16	85.5	6.7	860	1	SYL_ECO57
17	85.5	6.7	860	1	SYL_ECOL6
18	85.5	6.7	860	1	SYL_ECOL1
19	85.5	6.7	886	1	ITH3_MESAU
20	85	6.7	461	1	PBSC_CVAPA
21	84.5	6.6	501	1	AMPA_WIGBR
22	84.5	6.6	613	1	PEPF_MYCPU
23	84.5	6.6	1603	1	V174_CAEEL
24	84	6.6	439	1	SV62_DISOM
25	84	6.6	504	1	YC03_KLEPN
26	84	6.6	752	1	NECI_RAT
27	83.5	6.6	860	1	SYL_YERPE
28	83.5	6.6	1164	1	BAG_STRAG
29	82	6.4	336	1	YD48_METJA
30	82	6.4	353	1	HIS7_BUCAL
31	82	6.4	366	1	SE17_HUMAN
32	82	6.4	774	1	AMY2_SCHPO
33	81.5	6.4	589	1	SYD_HABDU
					P42212 aequorea vi
					P43827 haemophilus
					Q9pb98 xylella fas
					Q828h5 salmonella
					Q82g26 salmonella
					Q87c65 xylella fas
					Q8ehp4 shewanella
					Q88900 rattus norv
					P50859 candida gla
					P13846 heron hepat
					P44935 haemophilus
					P46024 haemophilus
					P19424 bacillus sp
					Q9pp92 campylobact
					O32629 haemophilus
					Q8xbn8 escherichia
					Q8fiy9 escherichia
					P07813 escherichia
					P97280 mesocricetu
					P48104 cyanophora
					Q8d295 wigglewort
					Q98gp0 mycoplasma
					P18947 caenorhabdi
					P24506 discopyge o
					Q48449 klebsiella
					P28840 rattus norv
					Q8zdf8 yersinia pe
					P27951 streptococ
					Q58743 methanococ
					P57203 buchnera sp
					Q8wt86 homo sapien
					O42918 schizosacch
					Q7vnf0 haemophilus

RESULT 1
GFP_AEQVI
ID GFP_AEQVI STANDARD; PRT; 238 AA.
AC P42212; Q17104; Q27903;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92175527; PubMed=1347277;
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA Cormier M.J.;
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein.";
RL Gene 111:229-233(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94185810; PubMed=8137953;
RA Inouye S., Tsuji F.I.;
RT "Aequorea green fluorescent protein. Expression of the gene and
RT fluorescence characteristics of the recombinant protein.";
RL FEBS Lett. 341:277-280(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
RT protein by modification of its codon usage.";
RL Plant Mol. Biol. 33:989-999(1997).
RN [4]
RP CHROMOPHORE.
RX MEDLINE=93192221; PubMed=8448132;
RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;
RT "Chemical structure of the hexapeptide chromophore of the Aequorea
RT green-fluorescent protein.";
RL Biochemistry 32:1212-1218(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=96355665; PubMed=8703075;
RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,
RA Remington S.J.;
RT "Crystal structure of the Aequorea victoria green fluorescent
RT protein.";
RL Science 273:1392-1395(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=98294543; PubMed=9631087;
RA Yang F., Moss L.G., Phillips G.N. Jr.;
RT "The molecular structure of green fluorescent protein.";
RL Nat. Biotechnol. 14:1246-1251(1996).

ALIGNMENTS

[7]
X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
MEDLINE-98455509; PubMed-9782051;
Wachter R.M., Eislinger M.A., Kallio K., Hanson G.T., Remington S.J.;
"Structural basis of spectral shifts in the yellow-emission variants
of green fluorescent protein.";
Structure 6:1267-1277(1998).
[8]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
MEDLINE-99238303; PubMed-10220315;
Eislinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
"Structural and spectral response of green fluorescent protein
variants to changes in pH.";
Biochemistry 38:5296-5301(1999).
- FUNCTION: Energy-transfer acceptor. Its role is to transduce the
blue chemiluminescence of the protein aequorin into green
fluorescent light by energy transfer. Fluoresces in vivo upon
receiving energy from the Ca(2+)-activated photoprotein aequorin.
Absorbs light maximally at 395 nm and exhibits a smaller
absorbance peak at 470 nm. The fluorescence emission spectrum
peaks at 509 nm with a shoulder at 540 nm.
- SUBUNIT: Monomer.
- TISSUE SPECIFICITY: Photocytes.
- PTM: Contains a covalently attached chromophore, which is composed
of modified amino acid residues. The chromophore is formed upon
cyclization of the residues Ser-dehydroTyr-Gly.
- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
chimeric proteins of GFP linked to other proteins where it
functions as a fluorescent protein tag. GFP tolerates N- and C-
terminal fusion to a broad variety of proteins. It has been
expressed in bacteria, yeast, slime mold, plants, Drosophila,
zebrafish, and in mammalian cells. As a noninvasive fluorescent
marker in living cells, it allows for a wide range of applications
where it may function as a cell lineage tracer, reporter of gene
expression, or as a measure of protein-protein interactions.
- DATABASE: NAMB-Protein Spotlight;
NOTE- Issue 11 of June 2001;
WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M62654; AAA27722.1; -;
EMBL; M62653; AAA27721.1; -;
EMBL; U29345; AAA58246.1; -;
EMBL; X96419; CAA65278.1; -;
PIR; J30692; JQ1514.
DR PDB; 1B9C; 17-NOV-00.
DR PDB; 1BFP; 07-JUL-97.
DR PDB; 1CAF; 14-JUN-00.
DR PDB; 1EWA; 08-NOV-96.
DR PDB; 1EMB; 16-JUN-97.
DR PDB; 1EMC; 20-AUG-97.
DR PDB; 1EME; 20-AUG-97.
DR PDB; 1EMF; 20-AUG-97.
DR PDB; 1EMG; 12-MAY-99.
DR PDB; 1EMK; 20-AUG-97.
DR PDB; 1EML; 20-AUG-97.
DR PDB; 1EMW; 20-AUG-97.
DR PDB; 1F09; 17-NOV-00.
DR PDB; 1F0B; 17-NOV-00.
DR PDB; 1GEL; 11-JAN-97.
DR PDB; 1H03; 15-JAN-02.
DR PDB; 1HUY; 04-JUL-01.
DR PDB; 1JBY; 07-JAN-03.
DR PDB; 1JBZ; 07-JAN-03.
DR PDB; 1KP5; 28-AUG-02.
DR PDB; 1KVP; 10-APR-02.

DR PDB; 1KVR; 10-APR-02.
DR PDB; 1KYS; 10-APR-02.
DR PDB; 1YFP; 28-OCT-98.
DR PDB; 2EMD; 20-AUG-97.
DR PDB; 2EMN; 20-AUG-97.
DR PDB; 2EMO; 20-AUG-97.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PD01229; GFLUORESCENT.
DR ProDom; PD013756; Green fl protein; 1.
KW Luminescence; 3D-structure.
FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).
FT MOD_RES 66 66 2,3-DIDEHYDROTYROSINE.
FT VARIANT 100 100 F -> Y.
FT VARIANT 108 108 T -> S.
FT VARIANT 141 141 L -> M.
FT VARIANT 219 219 V -> I.
FT CONFLICT 2 2 S -> G (IN REF. 3).
FT CONFLICT 25 25 H -> Q (IN REF. 2).
FT CONFLICT 80 80 Q -> R (IN REF. 3).
FT CONFLICT 157 157 Q -> P (IN REF. 2).
FT CONFLICT 172 172 E -> K (IN REF. 2).
FT HELIX 4 8
FT STRAND 12 22
FT TURN 23 24
FT STRAND 25 36
FT TURN 37 40
FT STRAND 41 48
FT TURN 49 50
FT HELIX 57 60
FT TURN 61 63
FT HELIX 69 71
FT STRAND 73 73
FT HELIX 76 81
FT HELIX 83 86
FT TURN 87 90
FT STRAND 92 100
FT TURN 101 102
FT STRAND 105 115
FT TURN 116 117
FT STRAND 118 128
FT TURN 132 133
FT TURN 135 139
FT STRAND 141 141
FT STRAND 148 155
FT TURN 156 159
FT STRAND 160 171
FT TURN 172 173
FT STRAND 176 187
FT STRAND 199 208
FT TURN 211 212
FT STRAND 217 227
SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21BFB6E05 CRC64;
Query Match 97.5%; Score 1241; DB 1; Length 238;
Best Local Similarity 97.9%; Pred. No. 3.3e-95;
Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTLPFTCTTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTLPFTCTTGKLPVWPPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHLEINYNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGIDFKEDGNILGHLEINYNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQQNTPIGDGPVLLPDNHYLTSQSALSKDPNEKDHMVLLGFVTAAGITLGNDELYK 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Db 181 HQQNTPTGDPVLLPDNHYLSTQSALSKDPNKRDRHVVLLFVTAAGITHGMDELYK 238

RESULT 2

ID	SYL_HAEIN	STANDARD	PRT	861 AA.
AC	P43527			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).			
GN	LEUS OR HI0921			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Haemophilus.			
NCBI_TaxID=727;				
SEQUENCE FROM N.A.				
STRAIN=rd / KW20 / ATCC 51907;				
MEDLINE=95350630; PubMed=7542800;				
Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,				
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,				
McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,				
Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,				
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,				
Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,				
Fine L.D., Fritchman J.L., Geoghagen N.S.M.,				
Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,				
Venter J.C.;				
"Whole-genome random sequencing and assembly of Haemophilus influenzae				
rd.";				
Science 269:496-512(1995).				
-!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +				
diphosphate + L-leucyl-tRNA(Leu).				
-!- SUBUNIT: Monomer (By similarity).				
-!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.				
This SWISS-PROT entry is copyright. It is produced through a collaboration				
between the Swiss Institute of Bioinformatics and the EMBL outstation -				
the European Bioinformatics Institute. There are no restrictions on its				
use by non-profit institutions as long as its content is in no way				
modified and this statement is not removed. Usage by and for commercial				
entities requires a license agreement (See http://www.isb-sib.ch/announcement/				
or send an email to license@isb-sib.ch).				
EMBL; U32774; AAC22581.1; -				
PIR; H64102; H64102.				
TIGR; HI0921; -				
HAMAP; MF 00049; -; 1.				
InterPro; IPR002302; Leu-trNASyntla.				
InterPro; IPR002300; tRNA-synt_1a.				
InterPro; IPR001412; tRNA-synt_1.				
InterPro; IPR009008; ValRS_ILERS_edit.				
PRINTS; PF00133; tRNA-synt 1; 1.				
PROSITE; PRO0985; TRNASYNTHLEU.				
TIGRFAMs; TIGR00396; leuS bact; 1.				
PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.				
Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;				
Complete proteome.				
SITE 42 52				
"HIGH" REGION.				
SITE 619 623				
"KMSKS" REGION.				
BINDING 622 623				
ATP (BY SIMILARITY).				
SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;				
Query Match				
Best Local Similarity 7.6%; Score 96.5; DB 1; Length 861;				
Matches 24.6%; Pred. No. 2.6;				
Conservative 26; Mismatches 59; Gaps 11;				
Indels 59; Gaps 11;				
50 TTGKLPVWPVPTLVLLSGVQCFSRYPDHMKQHDFFKSAMPEGVQVQRTIFFKD----- 103				
314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKYSLPKIQVIAPLA 364				

Db 181 HQQNTPTGDPVLLPDHNYLSTQSALSKDPNKRDRHWLVLEFVTAAAGITHGMDELYK 238

RESULT 2
SYL_HAEIN STANDARD; PRT; 861 AA.
ID AC P4327;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP + diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licens@isb-sib.ch).
CC EMBL; U32774; AAC22581.1; .
CC PIR; H64102; H64102.
CC TIGR; H10921; .
CC HAMAP; MF 00049; -. 1.
CC InterPro; IPR002302; Leu-trNASyntla.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_ILERS_edit.
CC Pfam; PF00133; tRNA-synt 1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFams; TIGR00396; leus bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
FT SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;

Query Match 7.6%; Score 96.5; DB 1; Length 861;
Best Local Similarity 24.6%; Pred. No. 2.6;
Matches 47; Conservative 26; Mismatches 59; Indels 59; Gaps 11;

Cc 50 TTGKLVPWPPTLVTLSSGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKD----- 103
Cc 314 TGDKLP-LP-WANFVLHMHYGTGAVMVAHPAH-DQRDF-----EFAQKYSLPIKQVIAPLA 364

Qy 104 DGYKTRAEVKFECDTLVNRLELGKIDPKEDGNLTGHKLENNYNSHNVIWADK-QKNGI 163
Db 365 DEEIDLTKQAFVEHGKLVNSDEFDGKNF--DGAENG-----IADKLEKLGV 408
Qy 163 ---KVNFKIRH-----NIEDGSVOLADHYQONTPTGDPVLLPDHNYL- 202
Db 409 GKRVNVRLDWGVSQRQRYWCAPIMPLTLENGDVVPA-----PMEDLPILLPEVDVMD 461
Qy 203 STQSALSCKPN 213
Db 462 GVKSPINADPN 472
RESULT 3
SYL_XYLFA STANDARD; PRT; 879 AA.
ID AC Q9PBG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR XF2176.
OS Xylella fastidiosae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares R.G., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krueger J.E., Kurame E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.W., Ishihako M.H.,
RA Vallado H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosae";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP + diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licens@isb-sib.ch).
CC EMBL; AE004031; AAF84975.1; ALT_INIT.
CC HAMAP; MF 00049; -. 1.

```

DR InterPro; IPR002302; Leu-TRNASyntla.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_fiers_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA-TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99796 MW; 9FDCB99202919E CRC64;

Query Match 7.4%; Score 94.5; DB 1; Length 879;
Best Local Similarity 22.7%; Pred. No. 3.9;
Matches 45; Conservative 29; Mismatches 67; Indels 57; Gaps 10;

QY 30 TTGKLPVPWPTLVTLISYGVQCSRYPDHMKQHDFFKSAHPGYYVOERTIFFKDDGNY-- 107
Db 321 TNEQLPV-WVANFVLMAYGTGAVMAVPGHDQDQEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 138 -----KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVI 153
Db 378 STWEPDVRDWDYADKTR---EFE---LNSAEFDGLDYQDAPEVLAERPE----- 421
QY 154 MADKQKNG-IKNFKIRHNIEDGSVOLADHYQONTPI-----GGGVLLLPDN 199
Db 422 ---RQGRQRRVNYRLR---DWGVSQRVWGCPVIVYCTGCGAVPEVDQLPVILPEN 474
QY 200 -HYLSTQSALSADPNEKR 216
Db 475 VAFSGTGSPKIDPEWRK 492

RESULT 4
SYL_SALTY STANDARD; PRT; 860 AA.
AC Q8Z8H5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS O3 STY0699 OR T2219.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]_TAXID=601;
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RA MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.F., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852 (2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyouranni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337 (2003).

CC CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC CC diphosphate + L-leucyl-tRNA(Leu).
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; AL627267; CAD05125.1; -.
DR EMBL; AE016841; AAO69822.1; -.
DR HAMAP; MF_00049; -. 1.
DR InterPro; IPR002302; Leu-TRNASyntla.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_fiers_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA-TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 860 AA; 96940 MW; 2F95E480BBAB23C4 CRC64;

Query Match 7.3%; Score 92.5; DB 1; Length 860;
Best Local Similarity 23.9%; Pred. No. 5.5; 77; Indels 39; Gaps 7;
Matches 43; Conservative 21; Mismatches 77; Indels 39; Gaps 7;

QY 50 TTGKLPVPWPTLVTLISYGVQCSRYPDHMKQHDFFKSAHPGYYVOERTIFFKDDGNYKT 109
Db 314 TGEIIPV-WANFVLMAYGTGAVMAVPGH-DQD-YEFASKYGLTIKPVILAADSGSEPD 370
QY 110 RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVIYINADKQKNGIKVNFIR 169
Db 371 SEQALTEKGVLFNSGFEFDGLAFNAFNAIDKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NIEDGSVOLADHYQONTPIGDDGVLLPDNHYI-STQSALSADP 212
Db 419 DWGVSQRVWGAPIPWVTLDEGTV-----LPTPEDQLPVILPEVDVMDGITSPKADP 471

RESULT 5
SYL_SALTY STANDARD; PRT; 860 AA.
AC Q8ZQZ6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR STM0648.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]_TAXID=602;
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).

```



```

RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblum T.V., Smith H.O., Venter J.C., Nealeon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to Class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE015561; AAN54244.1; -.
DR HAMAP; MF 00049; -.
DR TIGR; SO1174; -.
DR InterPro; IPR002302; Leu-tRNA-synt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_1leRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA TRNA_LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 618 622 "KMSKS" REGION.
FT BINDING 621 621 ATP (BY SIMILARITY).
SQ SEQUENCE 859 AA; 96827 MW; C78D6209DFB6CA17 CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 859;
Best Local Similarity 24.4%; Pred. No. 9.7;
Matches 51; Conservative 29; Mismatches 95; Indels 37; Gaps 9;

QY 18 ELDGVDNGHKFSVSGEGE-----GDATYGLTKLTKFICTGKLPVPWPTLVTLTSSGVQ 70
DB 277 ELAAFTDECKNSTTSAELATMEKRGVAT-GLYAIHPI--TGKQVPIANFVLMNYGTG 333
QY 71 CFSRYPDH-MKQHDFFKSPAMPEGVQERTIFFKDDGNKYKTRAEVKFEGDPLVNRIELKGI 129
DB 334 AVMSVPGHDQRDYEFAK----KVLPIEAIVKPAEGDLDISEAAYTEKGLFNSGEGDGL 389
QY 130 DKEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRH-----NIEDGSGVOLA 180
DB 390 DFDGAFNVIANKL-----VAGKGRK--QVNYRLRDMGVSQRQYWGAPIPMVTILA 437
QY 131 DHYQONTPTIGDGPVLLPDNHYL-STQSALSKD 211
DB 438 DGTVITPDELPLVPEDVMDGIQSPIKAD 469

RESULT 8
GRBE RAT
ID -GRBE RAT STANDARD; PRT; 538 AA.
AC O88970;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE GN Growth factor receptor-bound protein 14 (GRB14 adapter protein).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=98421528; PubMed=9748281;
RA Kasus-Jacobi A., Perdereau D., Auzan C., Clauser E., van Obberghen E.,
RA Mauvais-Jarvis F., Girard J., Burnol A.-F.;
RT "Identification of the rat adapter Grb14 as an inhibitor of insulin
RT actions.";
RL J. Biol. Chem. 273:26026-26035(1998).
CC -1- FUNCTION: Interacts with the cytoplasmic domain of the
CC autophosphorylated insulin receptor which is then inhibited. The
CC interaction is mediated by the SH2 domain.
CC -1- SUBUNIT: Binds to the ankyrin repeat region of TNKL via its N-
CC terminus (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC endosomes (By similarity).
CC -1- PTM: Phosphorylated on serine residues (By similarity).
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 1 Ras-associating domain.
CC -1- SIMILARITY: Contains 1 SH2 domain.
CC -1- SIMILARITY: Belongs to the GRB7/10/14 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF076619; AAC61478.1; -.
DR HSSP; P35235; IAYA.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000159; RA_domain.
DR InterPro; IPR000980; SH2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00788; RA; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00314; RA; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50200; RA; 1.
DR PROSITE; PS50001; SH2; 1.
KW SH2 domain; Phosphorylation.
FT DOMAIN 104 190 RAS-ASSOCIATING.
FT DOMAIN 232 340 PH.
FT DOMAIN 437 533 SH2.
SQ SEQUENCE 538 AA; 60592 MW; CEB9C9037E7868EEF CRC64;

Query Match 7.0%; Score 88.5; DB 1; Length 538;
Best Local Similarity 19.4%; Pred. No. 6.7;
Matches 41; Conservative 33; Mismatches 70; Indels 67; Gaps 9;

QY 41 GKLTAKFICT-TGKLPVPWPTLVTLTSSGVQCFSRV--PDHMKQHDFFKSPAMPEGVQSR 97
DB 313 GPRDLKMLCAEEDQSRMCWVTAILRLKYGMQLYQNYMHPQSARSACSSQSVSPMRSVSN 372
QY 98 TIFPKDDGNKYKTRAEVKFEGDPLVNRIELKGIIDFKEDG-----NIL 138
DB 373 SLVAMDFSGQKTRV-IDNPTEALSVAVE-EGLAWRKKGCLRLGNHGSPTAPSSAVNMA 430
QY 139 GHKLEYNNSH-----NVYIMADKQK-----GIKVPF----- 166
DB 431 LHRSQFWFHHRISRDQAQOLITROGPFVDGVFLVRDSQGNPRTFVLMSHSGQKIKHFILP 490

```


Best Local Similarity 21.8%; Pred. No. 9.8;
Matches 45; Conservative 32; Mismatches 80; Indels 49; Gaps 9

QY 167 -----KIRNIEDGS-----VOLADHYQON 186
DB 491 VEDDGEVFHTLDDGHTKFTDLQLVFEYQLN 521

RESULT 9
CP51_CANGA STANDARD; PRT; 533 AA.
ID CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPL1) (P450-IIA1) (Sterol 14-alpha-demethylase) (lanosterol 14-alpha demethylase) (P450-14DM).
DE ERG11 OR CYP51.
GN Candida glabrata (Yeast) (Torulopsis glabrata).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=2001-L5;
RC MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E., Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell viability, cell growth, sterol composition, and antifungal susceptibility.";
RT Antimicrob. Agents Chemother. 39:2708-2717(1995).
RL [2]
SEQUENCE OF 60-473 FROM N.A.
RP STRAIN=ATCC 2001;
RC MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J., Rossier M.;
RT "Rapid detection and identification of Candida albicans and Torulopsis (Candida) glabrata in clinical specimens by species-specific nested PCR amplification of a cytochrome P-450 lanosterol-alpha-demethylase (L1A1) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -1- FUNCTION: Catalyzes C14-demethylation of lanosterol which is critical for ergosterol biosynthesis. It transforms lanosterol into 4,4'-dimethyl cholesterol.8,14,24-triene-3-beta-ol (By similarity).
CC -1- CATALYTIC ACTIVITY: Obtusifolliol + 3 O(2) + 3 NADPH = 4-alpha-methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3 H2O.
CC -1- PATHWAY: Ergosterol biosynthesis.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; L40389; AAB02329.1; -;
DR EMBL; S75389; AAB32679.1; -;
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PR00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
FT Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507B6EF7 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 19.7%; Pred. No. 16;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;

QY 58 WPTLVTLISYGVQCFSRYPDHMKQH-----DFFKSAPEGVVQERT-----IPFKDGNKY 108
DB 139 WPKGISYLPVHSGVKPYPEQQNHESLVNDYLNKLFAGILYKRVSKHLVTFK--GPYF 196

Best Local Similarity 21.8%; Pred. No. 9.8;
Matches 45; Conservative 32; Mismatches 80; Indels 49; Gaps 9

QY 25 GHKFSVS---GEGEGDATYKGLTKLFTCTTGKLPVWPMTLVTLISYGVQCFSRYPDH--M 79
DB 109 GHEFIFNAKLADVSAEAAYSHLT-----TPV-----FGKGVYDCPNHRLM 149

QY 80 QHDFFKSAM-PEGYV-----QERTIFFKDDGNYKTRAEVKFEFGDTLVNRIELKGIDF 131
DB 150 EQKFEVKGALTKEAFVRYVPIIAEEIYKYPNSKNFKINENNSGIVDMVMSQPEM--TIF 207

QY 132 KEDGNILGHKLEYNVNNHNVIMADKOKNGIKVNFKIRHNIEDSGSVLADHYQONTPEIGD 191
DB 208 TASRSLLEKEMRDKLDTDFAVLVSOLDKGTPTINF-VFPNLEPLEHYKRDHAQQAIS--- 263

QY 192 GPVLLPDNHYLSTQSALSQKDFNEKD 217
DB 264 -----GTYSMLIKERREKND 278

RESULT 10
DPOL HPBHE STANDARD; PRT; 788 AA.
ID DPOL HPBHE STANDARD; PRT; 788 AA.
AC P13846;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P protein [includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].
GN Heron hepatitis b virus.
OS Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=28300;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88333160; PubMed=3418788;
RA Sprengel R., Kalaeta E.F., Will H.;
RT "Isolation and characterization of a hepatitis B virus endemic in herons.";
RL J. Virol. 62:3832-3839(1988).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate + [DNA](N).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-phosphomonoester.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; M22056; AAA45738.1; -;
DR PIR; A30082; JDVLHLLH.
DR InterPro; IPR001462; DNAPol_viral_C.
DR InterPro; IPR000201; DNAPol_viral_N.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00336; DNA_pol_viral_C; 1.
DR Pfam; PF00242; DNA_pol_viral_N; 1.
DR Pfam; PF00078; rvt; 1.
DR ProDom; PD000814; DNAPol_viral_C; 1.
KW Transferase; RNA-directed DNA polymerase; DNA-directed DNA polymerase; Hydroxylase; Nuclease; Endonuclease; DNA replication; DNA-binding.
SQ SEQUENCE 788 AA; 90070 MW; FB44F38F75EADFA4 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 788;
Best Local Similarity 19.7%; Pred. No. 16;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;

QY 58 WPTLVTLISYGVQCFSRYPDHMKQH-----DFFKSAPEGVVQERT-----IPFKDGNKY 108
DB 139 WPKGISYLPVHSGVKPYPEQQNHESLVNDYLNKLFAGILYKRVSKHLVTFK--GPYF 196

```
109 T-----BAEVKFGDTLVNRILKGIQKEDGNILGHKLEYNVSHNVYIM 154
197 TWEQKHLVPOOHGAYSKINDROESRRRIITATSRKNDSSRI-----FGAHN----- 245
155 ADKQKNGIKVNFIRNIEDGVSQVLADHYQ-----QNTPIGDGCVLL-PDNHYL 202
246 -----NGRKISY---HSTRDGSRLSRTSDPTSRGALAGDSTPIGPGSTAHPSTHIV 297
203 STQ-----SALSQDPNEKR 216
298 DRRRQKQGVLOAISEPSETR 320

RESULT 11
D152_HAEIN STANDARD; PRT; 795 AA.
ID D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN HI0917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RG / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Keniake A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32773; AAC22575.1; -
CC TIGR; HI0917; -
CC InterPro; IPR000184; Bac_surfAg_D15.
CC Pfam; PF01103; Bac_surface_Ag; 1.
CC Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
FT SEQUENCE 795 AA; 87478 MW; B85691FC2256ED44 CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 795;
Best Local Similarity 21.9%; Pred. No. 16;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;
QY 65 LSYGVQCFSRYPDHMKQHDF-----FKSAMPEGYVOE-----RTI 99
Db 427 IGYGTSGISYQASVKQDNFLGTGAAGVSTAGTNDYGTSNVLYGTFPTKDGVS LGNV 486
QY 100 FFXDDGNKYTRAEVKEFGDTLVNRILKGIQKEDGNI---LGH-----KLEYNYS 148
```

```
487 FFENVNDSKSDTSSNYKRTTYGSNVTL-GFPVNNNSYTVGLGHTYKNISNFALEYN--- 542
149 HNVYIMADKQK-NGIKVNFIRNIEDGVSQVLADHYQQ-----NTPIGDGPVLL 196
543 RNLTYQSMFKNGIKTN-----DFDFSGFNYSNLSNRGYFPTKGVKASLG-GRVTI 593
197 P--DNHYLSTQALSQDPNEKRDMVLLGVTAAGITLG 233
594 PGSDNKYKLSADVQGFYFLDRDLHVLWVSAKASAGYANG 632

RESULT 12
D151_HAEIN STANDARD; PRT; 797 AA.
ID D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loomore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Winna / Serotype B, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U13961; AAA85645.1; -
CC DR EMBL; U60832; AAB61974.1; -
CC DR EMBL; U60833; AAB61976.1; -
CC DR PIR; JC4078; JC4078.
CC InterPro; IPR000184; Bac_surfAg_D15.
CC Pfam; PF01103; Bac_surface_Ag; 1.
CC Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
FT SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF1B CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 797;
Best Local Similarity 21.9%; Pred. No. 16;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;
QY 65 LSYGVQCFSRYPDHMKQHDF-----FKSAMPEGYVOE-----RTI 99
Db 427 IGYGTSGISYQASVKQDNFLGTGAAGVSTAGTNDYGTSNVLYGTFPTKDGVS LGNV 486
```

```
QY 100 FFKDDGNYKTRAEVKEGDTLVNRIELKGIIDFKEDGNI-----LGH-----KLEYNYS 148
Db 487 FFENYDNSKDTSSNYKRTTYGNSVTL-GPPVNNNSYVGLGHTYNTKISNFALEYN--- 542
QY 149 HNVYIMADKOK-NGIKVNFKIRHNIEDGVSQVLADHYQQ-----NTPIGDGPVLL 196
Db 543 RNDYIOSMKFKGNGIKTN-----DFDFSGWNNYNSLNRGVPPTKGKVASLG-GRVTI 593
QY 197 P--DNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLG 233
Db 594 PGSDNKKYKLSADVQGFYPLDRDHLWVWSAKASAGYANG 632

RESULT 13
GUN_BACS6
ID GUN_BACS6 STANDARD; PRT; 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Alkaline cellulase).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1415;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=91037937; PubMed=2230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT "Molecular cloning and nucleotide sequence of a gene for alkaline
cellulase from Bacillus sp. KSM-635."
RL J. Gen. Microbiol. 136:1327-1334(1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
hydrolases).
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27420; AAA22304.1; -.
DR F1R; S29043; S29043.
DR PDB; 1G01; 31-DEC-02.
DR PDB; 1G0C; 31-DEC-02.
DR InterPro; IPR005086; CBM_17_28.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR001119; SLH.
DR Pfam; PF03424; CBM_17_28; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW Cellulose degradation; Hydrolyase; Glycosidase; Signal; Repeat;
KW 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 941 ENDOGLUCANASE.
FT DOMAIN 40 99 SLH 1.
FT DOMAIN 100 151 SLH 2.
FT DOMAIN 152 225 SLH 3.
FT ACT_SITE 373 373 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 485 485 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 941 AA; 104628 MW; BEA2AC3B169BFADA CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 941;
Best Local Similarity 20.3%; Pred. No. 19;
Matches 49; Conservative 34; Mismatches 61; Indels 97; Gaps 11;
```

```
QY 16 LVELDGDVNGHKFVSSEGEEDATYKGLTLKFICTTGKLPVWPMTLVLLSYGVQCFERY 75
Db 240 LVELNG-----QLTLAGE---DGT-----PV-----QLRGMSTHGLQWFG-- 271
QY 76 PDHMKQHDFFKSGAMPGEYVOERTIFPKDGNQYKTRAEVKEGDTLVNRIELKGIIDFKEDG 135
Db 272 -EIVNENAFVLSNDWGSNMIRLAMYIGENVATNPEVK---DLVYEGIELA----- 319
QY 136 NILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGVS---OLADHYQQNTPIGD 191
Db 320 -----FEHDMYIVDWH---VHAPGDPADVYSGAYDFFFEIADHYKDH----- 360
QY 192 GPVLLPDNHYLSTQSALSADPN-----EKRDMHMLV 222
Db 361 -----PKHYIITWELANEPSPNNGGPGLTNDEKGWEAVEPIVEMLRKGDNMILV 415
QY 223 G 223
Db 416 G 416

RESULT 14
TRNA_CAMJE
ID TRNA_CAMJE STANDARD; PRT; 357 AA.
AC Q9PP92;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (tRNA (M-5-U54)-
methyltransferase) (RUMT).
DE TRMA OR CU0831C.
GN Campylobacter jejuni.
OS Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -!- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
54 (M-5-U54) in all tRNA (By similarity).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
homocysteine + tRNA containing thymine.
CC -!- SIMILARITY: Belongs to the RNA MSU methyltransferase family. TrmA
subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL139076; CAB73096.1; -.
DR F1R; G81355; G81355.
DR HAMAP; MF_01011; -.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR001566; TrmA.
DR PROSITE; PS01230; TRMA_1; 1.
DR PROSITE; PS01231; TRMA_2; FALSE_NEG.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY
SIMILARITY).
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds
(without alignments)
2458.984 Million cell updates/sec

Title: US-09-887-784-641
Perfect score: 1273
Sequence: 1 MVSKEELFTGVVILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_25:*
 - 1: sp archaea:*
 - 2: sp bacteria:*
 - 3: sp fungi:*
 - 4: sp human:*
 - 5: sp invertebrate:*
 - 6: sp mammal:*
 - 7: sp mhc:*
 - 8: sp organelle:*
 - 9: sp phase:*
 - 10: sp plant:*
 - 11: sp rodent:*
 - 12: sp virus:*
 - 13: sp vertebrate:*
 - 14: sp unclassified:*
 - 15: sp virus:*
 - 16: sp bacteriaph:*
 - 17: sp archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1237	97.2	238	Q8GHE2	Q8ghe2 azotobacter
2	1234	96.9	238	Q93125	Q93125 aequorea vi
3	1232	96.8	238	Q8GHE4	Q8ghe4 azomonas ag
4	1231	96.7	238	Q8GHE3	Q8ghe3 azotobacter
5	1199	94.2	238	Q17105	Q17105 aequorea vi
6	1184	93.0	238	Q17106	Q17106 aequorea vi
7	1079	84.8	238	Q8WTC6	Q8wtc6 aequorea ma
8	1075	84.4	238	Q8WP95	Q8wp95 aequorea ma
9	1071	84.1	238	Q8WTC4	Q8wtc4 aequorea ma
10	1069	84.0	238	Q8WTD0	Q8wtcd0 aequorea ma
11	1068	83.9	238	Q8WTC8	Q8wtc8 aequorea ma
12	1068	83.9	238	Q8WTC9	Q8wtc9 aequorea ma
13	1066	83.7	238	Q8WTC7	Q8wtc7 aequorea ma
14	1064	83.6	238	Q8WTC5	Q8wtc5 aequorea ma
15	251.5	19.8	225	Q95UA7	Q95ua7 montastraea
16	251.5	19.8	225	Q7Z0W5	Q7z0w5 montastraea

17	246	19.3	225	5	Q963F5	Q963f5 montastraea
18	243.5	19.1	225	5	Q7Z0W9	Q7z0w9 montastraea
19	243.5	19.1	236	5	Q8T6U0	Q8t6u0 dendronephc
20	239	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyll
21	237.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q8T5F1	Q8t5f1 montastraea
23	232	18.2	225	5	Q7Z0W4	Q7z0w4 montastraea
24	231	18.1	224	5	Q8MU48	Q8mu48 montastraea
25	215	16.9	227	5	Q7Z0W6	Q7z0w6 montastraea
26	215	16.9	234	5	Q7Z0W7	Q7z0w7 montastraea
27	211.5	16.6	259	5	Q8WMA2	Q8wma2 agaricia fr
28	211	16.6	239	5	Q8WMA1	Q8wma1 agaricia ag
29	210	16.5	234	5	Q8T5F2	Q8t5f2 montastraea
30	210	16.5	234	5	Q8MU47	Q8mu47 montastraea
31	209.5	16.5	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	207	16.3	227	5	Q962P9	Q962p9 montastraea
33	207	16.3	227	5	Q7Z0W8	Q7z0w8 montastraea
34	203	15.9	221	5	Q95P04	Q95p04 goniorpara t
35	203	15.9	227	5	Q95V70	Q95v70 montastraea
36	202.5	15.9	232	5	Q9GP15	Q9gpi15 anemonia su
37	202.5	15.9	238	5	Q9BLV9	Q9bly9 renilla mue
38	200.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
39	198.5	15.6	214	5	Q861V7	Q861v7 meandrina m
40	198.5	15.6	232	5	Q9GZ28	Q9gz28 anemonia su
41	197.5	15.5	214	5	Q86LV8	Q86lv8 meandrina m
42	197.5	15.5	222	5	Q7Z168	Q7z168 cerianthus
43	197.5	15.5	225	5	Q8T6T9	Q8t6t9 radianthus
44	197	15.5	235	5	Q8T5F0	Q8t5f0 scolymia cu
45	196.5	15.4	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	PRELIMINARY;	PRT;	238 AA.
ID	Q8GHE2		
AC	Q8GHE2;		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Green fluorescence protein.		
GN	2289GFP.		
OS	Azotobacter vinelandii.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
OC	Pseudomonadaceae; Azotobacter.		
OX	NCBI_taxid=354;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=DSM2289;		
RA	Koranyi P., Berenyi M., Burg K.;		
RT	Occurrence of green fluorescence protein in diazotrophic bacteria		
RT	Azomonas and Azotobacter.;		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF324408; AAN86140.1; -.		
DR	GO; GO:0006091; P:energy pathways; IEA.		
DR	InterPro; IPR009017; GFP-like.		
DR	InterPro; IPR000786; Green_fl_protein.		
DR	Pfam; PF01353; GFP; 1.		
DR	PRINTS; PR01229; GFP_LIKE.		
DR	PRODOM; PD013756; Green fl protein; 1.		
SQ	SEQUENCE 238 AA; 26914 MW; P84840F1F9064018 CRC64;		

Query Match	97.2%;	Score 1237;	DB 2;	Length 238;
Best Local Similarity	97.5%;	Pred. No. 1.5e-96;		
Matches 232;	Conservative	2;	Mismatches 4;	Indels 0;
Gaps	0;			
Qy	2	VSKGEELFTGVVILVELDGVNKHFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL	61	
Db	1	MSKGEELFTGVVILVELDGVNKHFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL	60	
Qy	62	VTLLSYGVQCFSPYDPMKQHDFFKSPMEGYVOERTIPFKDGNKYKTRAEVPEGDTLV	121	

```
Db 61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHLEYNINSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHLEYNINSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 2
Q93125 ID Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,
RA Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
DR SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 96.9%; Score 1234; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 2.7e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGNVGHKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPILVELDGNVGHKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL 60

QY 52 VTLLSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 51 VTTFGYGVQCFARYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHLEYNINSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHLEYNINSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 3
Q8GHE4 ID Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
```

```
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324405; AAN86137.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
DR SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.8%; Score 1232; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 4e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGNVGHKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPILVELDGNVGHKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHLEYNINSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHLEYNINSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 4
Q8GHE3 ID Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
DR SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.7%; Score 1231; DB 2; Length 238;
```

```
Best Local Similarity 97.1%; Pred. No. 4.9e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKRHDFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.2%; Score 1199; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 2.5e-93;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGQKFSVSGEGDATYKGLTLNFICTTGKLPVPWPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKMEYNYNHNVYIMADKPKNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 93.0%; Score 1184; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 4.6e-92;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGQKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKMEYNYNHNVYIMADKPKNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.O., Chen M.,
RA Li S.J., Xia N.S.;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
```

```
DR PRINTS; PRO1229; GFP, FLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 6B8FD75E88926903 CRC64;

Query Match      84.8%; Score 1079; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 3.4e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKLTIKFICTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVHGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSRYPDHMKOHDFPKSAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTFSYGIQCFARYPEHMKNDFFPKSAMPEGYQERTIFFDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHYVIMPDKANNGLKNVFKIRHNIEGGVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLLIPINHYLSTQTAISKDRNETRDHMLVFLFFSACGTHGMDELYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFP, 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.1%; Score 1071; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.6e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKLTIKFICTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPILVELDGDVHGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSRYPDHMKOHDFPKSAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLVGILCFARYPEHMKNDFFPKSAMPEGYQERTIFFDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHYVIMPDKANNGLKNVFKIRHNIEGGVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLLIPINHYLSTQTAISKDRNETRDHMLVFLFFSACGTHGMDELYK 238

RESULT 9
Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFP, 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.4%; Score 1071; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.6e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKLTIKFICTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPILVELDGDVHGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSRYPDHMKOHDFPKSAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLVGILCFARYPEHMKNDFFPKSAMPEGYQERTIFFDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHYVIMPDKANNGLKNVFKIRHNIEGGVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLLIPINHYLSTQTAISKDRNETRDHMLVFLFFSACGTHGMDELYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFP, 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.4%; Score 1075; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 7.4e-83;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKLTIKFICTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPVLIELDGDVHGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSRYPDHMKOHDFPKSAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTFSYGIQCFARYPEHMKNDFFPKSAMPEGYQERTIFFDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHYVIMPDKANNGLKNVFKIRHNIEGGVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLLIPINHYLSTQTAISKDRNETRDHMLVFLFFSACGTHGMDELYK 238
```


RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF435427; AAL33912.1; --
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP like.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFLUORESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 84.0%; Score 1069; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 2.4e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGNGHFKFSVSGEGDATYKGLTKLFICTTGKLPVPWPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVPWPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTLGYGIOCFARYPEHMKQNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY 122 NRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 NRLEKGMDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEGGVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSQTALSADKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HYQTNVPLGDGPVLLPINHYLSQTALSADKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 238

RESULT 11

Q8WTC8 PRELIMINARY; PRT; 238 AA.

AC Q8WTC8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435429; AAL33914.1; --
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP like.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFLUORESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 83.9%; Score 1068; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 2.9e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGNGHFKFSVSGEGDATYKGLTKLFICTTGKLPVPWPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVPWPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTLGYGIOCFARYPEHMKQNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 NRLEKGMDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEGGVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSQTALSADKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HYQTNVPLGDGPVLLPINHYLSQTALSADKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 238

RESULT 12

Q8WTC9 PRELIMINARY; PRT; 238 AA.

AC Q8WTC9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.

OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm162;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435428; AAL33913.1; --
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP like.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFLUORESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match 83.9%; Score 1068; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 2.9e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGNGHFKFSVSGEGDATYKGLTKLFICTTGKLPVPWPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVPWPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTLGYGIOCFARYPEHMKQNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY 122 NRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 NRLEKGMDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEGGVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSQTALSADKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HYQTNVPLGDGPVLLPINHYLSQTALSADKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 238

RESULT 13

Q8WTC7 PRELIMINARY; PRT; 238 AA.

AC Q8WTC7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;

RN	[1]	SEQUENCE FROM N.A.	
RP	STRAIN=GFPxm191uv;		
RA	Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,		
RA	Li S.J., Xia N.S.;		
RT	"Colorful mutants of green fluorescent protein from Aequorea		
RT	macrodactyla.";		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF435430; AAL33915.1; -		
DR	GO: GO:0006091; P:energy pathways; IEA.		
DR	InterPro; IPR009017; GFP like.		
DR	InterPro; IPR000786; Green_fl_protein.		
DR	Pfam; PF01353; GFP; 1.		
DR	PRINTS; PR01229; GFPLORESCENT.		
DR	ProDom; PD013756; Green fl protein; 1.		
SQ	SEQUENCE 238 AA; 27002 MW; BD5BA29822624C018 CRC64;		
Query Match 83.7%; Score 1066; DB 5; Length 238;			
Best Local Similarity 81.1%; Pred. No. 4.3e-82;			
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;			
QY	2	VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGGDATYGKLTFLKFCITGKLPVWPPTL 61	
DB	1	MSKGEELFTGIVPVLIELDGDVGHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 60	
QY	62	VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121	
DB	61	VTTLSYGILCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120	
QY	122	NRIELKGIDPKEGNILGHLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181	
DB	121	NRIELKGMDPKEDGNILGHLEYNFNHNYIIPDKANGLKYNFKIRHNIEGGVQLAD 180	
QY	182	HYQONTPIGDPVLLPDNHYLSTQSALSKDPNPKRDHMLVLLGFTVAAGITLGMDELYK 239	
DB	181	HYQTNVPLGDPVLPINHYLSYQTAISKDRNETRDHMLVLEFFSACGTHGMDELYK 238	
RESULT 14			
Q8WTC5	Q8WTC5	PRELIMINARY; PRT; 238 AA.	
AC	Q8WTC5;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Orange fluorescent protein.		
GN	GFP.		
OS	Aequorea macrodactyla.		
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;		
OC	Aequoreidae; Aequorea.		
OX	NCBI_Taxid=147615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=OPFxm;		
RA	Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,		
RA	Li S.J., Xia N.S.;		
RT	"Colorful mutants of green fluorescent protein from Aequorea		
RT	macrodactyla.";		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF435432; AAL33917.1; -		
DR	GO: GO:0006091; P:energy pathways; IEA.		
DR	InterPro; IPR009017; GFP like.		
DR	InterPro; IPR000786; Green_fl_protein.		
DR	Pfam; PF01353; GFP; 1.		
DR	PRINTS; PR01229; GFPLORESCENT.		
DR	ProDom; PD013756; Green fl protein; 1.		
SQ	SEQUENCE 238 AA; 27018 MW; 75521EA5534B573A CRC64;		
Query Match 83.6%; Score 1064; DB 5; Length 238;			
Best Local Similarity 81.9%; Pred. No. 6.3e-82;			
Matches 195; Conservative 17; Mismatches 26; Indels 0; Gaps 0;			
QY	2	VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGGDATYGKLTFLKFCITGKLPVWPPTL 61	

DB	1	MSKGEELFTGVVPILVELDGDVGHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 60	
QY	62	VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121	
DB	61	VTTLSYGILCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120	
QY	122	NRIELKGIDPKEGNILGHLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181	
DB	121	NRIELKGMDPKEDGNILGHLEYNFNHNYIIPDKANGLKYNFKIRHNIEGGVQLAD 180	
QY	182	HYQONTPIGDPVLLPDNHYLSTQSALSKDPNPKRDHMLVLLGFTVAAGITLGMDELYK 239	
DB	181	HYQTNVPLGDPVLPINHYLSYQTAISKDRNETRDHMLVLEFFSACGTHGMDELYK 238	
RESULT 15			
Q95UA7	Q95UA7	PRELIMINARY; PRT; 225 AA.	
ID	Q95UA7;		
AC	Q95UA7;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Cyan fluorescent protein (fragment).		
OS	Montastraea cavernosa (great star coral).		
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;		
OC	Favina; Faviidae; Montastraea.		
OX	NCBI_Taxid=63558;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Falkowski P.G., Sun Y.;		
RT	"Montastraea cavernosa fluorescent protein.";		
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AV056460; AAL17905.1; -		
DR	GO: GO:0006091; P:energy pathways; IEA.		
DR	InterPro; IPR009017; GFP like.		
DR	InterPro; IPR000786; Green_fl_protein.		
DR	Pfam; PF01353; GFP; 1.		
DR	PRINTS; PR01229; GFPLORESCENT.		
DR	ProDom; PD013756; Green fl protein; 1.		
FT	NON_TER		
SQ	SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;		
Query Match 19.8%; Score 251.5; DB 5; Length 225;			
Best Local Similarity 31.1%; Pred. No. 2.6e-13;			
Matches 64; Conservative 43; Mismatches 82; Indels 17; Gaps 7;			
QY	12	VPILVELDGDVNGHKFSVSGEGGDATYKLTFLKFCITGKLPVWPPTLVLSSGVQ 70	
DB	7	VMKIKLRMDGIVNGHKFMITGEGKPFEGTHTIILKVKEGGPPFPAYDILLTTAFQYGNR 66	
QY	71	CFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID 130	
DB	67	VFTKYPKDPD--DYFKQSPFEGYSWERSMTFFDQGVCTVTSIDIKLGDGCFYEIFYGVN 124	
QY	131	FKEDGNILGHK-LEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLADHYQONTPI 189	
DB	125	FPSSGPMQKTKLKEPSTENMYV----RDGVLGDVSVSTLLLEGD----KHRCNFRS 175	
QY	190	GDGP---VLIPDNHYLSTQ-SALSKD 211	
DB	176	TYGAKKGVLPEYHFDHRIEILSHD 201	
Search completed: June 21, 2004, 16:00:10			
Job time : 31.7778 secs			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds
(without alignments)
1433.395 Million cell updates/sec

Title: US-09-887-784-64V
Perfect score: 1273
Sequence: 1 MYSKGBELFTGVVILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*1: Genesecp1980s:*2: Genesecp1990s:*3: Genesecp2000s:*4: Genesecp2001s:*5: Genesecp2002s:*6: Genesecp2003as:*7: Genesecp2003bs:*8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1269	99.7	239	5 AAE17518	Aae17518 Enhanced
2	1269	99.7	363	6 ABR40352	Abr40352 Human ami
3	1269	99.7	893	4 AAG65781	Aag65781 Amino aci
4	1269	99.7	1132	4 AAG55782	Aag55782 Amino aci
5	1261	99.1	239	5 AAE17517	Aae17517 Enhanced
6	1258	98.8	239	3 AAB22882	Aab22882 Enhanced
7	1258	98.8	239	3 AAY54349	Aay54349 Amino aci
8	1258	98.8	239	3 AAY79584	Aay79584 EGFP sign
9	1258	98.8	239	4 AAB50804	Aab50804 Jellyfish
10	1258	98.8	239	4 AAB85900	Aab85900 A. victor
11	1258	98.8	239	4 AAB31171	Aab31171 Amino aci
12	1258	98.8	239	5 AAG66198	Aag66198 A. victor
13	1258	98.8	239	5 AAG94444	Abg94444 Protease
14	1258	98.8	239	5 AAE14599	Aae14599 Aequorea
15	1258	98.8	239	6 AAE34958	Aae34958 Aequorea
16	1258	98.8	239	6 AAG79829	Aag79829 Green flu
17	1258	98.8	239	6 ABR83616	Abr83616 Green flu
18	1258	98.8	239	6 ADA38074	Ada38074 Aequorea
19	1258	98.8	239	7 ABU63204	Abu63204 Aequorea
20	1258	98.8	239	7 ADC18358	Adc18358 EGFP (enh
21	1258	98.8	239	7 ABW00914	Abw00914 Aequorea
22	1258	98.8	239	7 ADE28570	Ade28570 Enhanced
23	1258	98.8	246	7 ABM79011	Abm79011 Enhanced
24	1258	98.8	248	5 AAG68319	Aag68319 Jellyfish
25	1258	98.8	259	5 AAU99804	Aau99804 Biomembra

26	1258	98.8	265	2 AAW97451	Aaw97451 Wild-type
27	1258	98.8	268	5 AAU99803	Aau99803 Biomembra
28	1258	98.8	270	5 AAU99802	Aau99802 Biomembra
29	1258	98.8	272	5 AAU99800	Aau99800 Biomembra
30	1258	98.8	273	5 AAU99801	Aau99801 Biomembra
31	1258	98.8	280	5 AAU99807	Aau99807 Biomembra
32	1258	98.8	281	3 AAY50142	Aay50142 Green flu
33	1258	98.8	281	3 AAB24252	Aab24252 EGFP-MODC
34	1258	98.8	281	5 AAU10888	Aau10888 EGFP-MODC
35	1258	98.8	286	7 ADE28562	Ade28562 EGFP/ hum
36	1258	98.8	289	7 ADE28564	Ade28564 EGFP/ hum
37	1258	98.8	290	7 ADE28568	Ade28568 EGFP/ hum
38	1258	98.8	290	7 ADE28566	Ade28566 EGFP/ hum
39	1258	98.8	294	3 AAB22860	Aab22860 GFP-DSVD-
40	1258	98.8	294	3 AAY79638	Aay79638 Caspase-3
41	1258	98.8	294	5 ABG94422	Abg94422 Recombina
42	1258	98.8	308	2 AAY42181	Aay42181 EGFP/DRM
43	1258	98.8	320	6 ABR83620	Abr83620 HUB1-GFP
44	1258	98.8	323	3 AAY54359	Aay54359 GFP mutan
45	1258	98.8	323	6 ABR83621	Abr83621 RUB1-GFP

ALIGNMENTS

RESULT 1
AAE17518
ID AAE17518 standard; protein; 239 AA.
XX
AC AAE17518;
DT XX
DT 22-APR-2002 (first entry)
XX
DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.
XX
KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
KW cellular function; genetic reporter; mutant; Stoke's shift; muten.
XX
OS Aequorea victoria.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 65 /note= "Wild type Phe substituted with Leu; This
FT corresponds to position 64 in the wild type protein"
FT
FT Misc-difference 223 /note= "Wild type Glu substituted with Gly; This
FT corresponds to position 222 in the wild type protein"
FT
FT WO200198338-A2.
PD 27-DEC-2001.
XX
XX 18-JUN-2001; 2001WO-EP006848.
PR 19-JUN-2000; 2000DK-00000953.
PR 20-JUN-2000; 2000US-0212681P.
PR 10-MAY-2001; 2001DK-00000739.
PR 10-MAY-2001; 2001US-0290170P.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX Bjorn SP, Pagliaro L, Thastrup O;
XX WPI; 2002-098224/13.
DR N-PSDB; AAD28163.
XX
PT Novel fluorescent protein in in vitro assay for measuring protein kinase
PT activity or dephosphorylation activity, or for measuring protein
PT redistribution, has a green fluorescent protein with F64L and E222G
PT mutation.
XX
PS Claim 9; Page 37; 41pp; English.

XX CC The invention relates to a fluorescent protein derived from green
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
CC at F64L and E222G has a bigger compared to other GFP's making it very
CC suitable for high throughput screening due to better resolution. The
CC fluorescent protein is useful in invitro assays for measuring protein
CC kinase activity or dephosphorylation activity, or for measuring protein
CC redistribution. The fluorescent protein is useful in studying cellular
CC functions in living cells; as protein tags in transgenic animals, living
CC and fixed cells; organelle tags, secretion marker and genetic reporter.
CC The fluorescent protein is also useful as a cell or organelle integrity
CC marker, a marker for changes in cell morphology, as transfection marker,
CC and as a marker to be used in combination with fluorescence activated
CC cell sorting (FACS). The novel proteins can also be used as reporters to
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
CC protein is also useful as markers in transcriptional and translational
CC fusions for performing transposon vector mutagenesis and as a reporter
CC for bacterial detection. Transposons encoding the fluorescent protein are
CC useful for screening promoters and for tagging plasmids and chromosomes.
CC The fluorescent protein engineered into the genome of a phage is useful
CC for designing diagnostic tool. The present sequence is a DNA encoding
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
XX SQ Sequence 239 AA;

Query Match 99.7%; Score 1269; DB 5; Length 239;
Best Local Similarity 99.6%; Pred. No. 9.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 2
ABR40352
ID ABR40352 standard; protein; 363 AA.
XX ABR40352;
AC ABR40352;
XX ABR40352;
DT 08-JUL-2003 (first entry)
XX Human amino acid sequence SEQ ID NO: 6.
XX Human; heterologous conjugate; intracellular protein.
XX Homo sapiens.
XX Aequoria victoria.
XX WO2003029827-A2.
XX 10-APR-2003.
XX 01-OCT-2002; 2002WO-DK000651.
XX 01-OCT-2001; 2001DK-00001433.
XX 11-OCT-2001; 2001US-0328896P.
XX (BIOL-) BIOLMAGE AS.
XX Terry BR, Nielsen SJ;

XX DR WPI: 2003-430211/40.
DR N-PSDB; ACC72604.
XX Novel cell for identifying modulators of protein interaction, contains a
PT first conjugate comprising anchor protein, second conjugate having type B
PT interactor protein and third conjugate with detectable group.
XX Disclosure; Page 112-113; 118pp; English.
XX The invention relates to a novel cell, comprising three heterologous
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that
CC specifically binds to an internal structure within the cell conjugated to
CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
CC type B conjugated to a first protein of interest, and a third HC (HC3)
CC comprising a second protein of interest conjugated to detectable group.
CC The cell is useful for detecting if a compound disrupts or induces the
CC interaction between two intracellular proteins. The cell is also useful
CC for screening compounds that modulate the interaction between two
CC intracellular proteins. The present sequence is used in the
CC exemplification of the invention
XX SQ Sequence 363 AA;

Query Match 99.7%; Score 1269; DB 6; Length 363;
Best Local Similarity 99.6%; Pred. No. 1.8e-122;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 3
AAG65781
ID AAG65781 standard; protein; 893 AA.
XX AAG65781;
AC AAG65781;
XX 07-JAN-2002 (first entry)
XX Amino acid sequence of HSPDB4A1-E222G fusion protein.
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX fusion protein.
XX Homo sapiens.
XX Aequoria victoria.
XX WO200179526-A2.
XX 25-OCT-2001.
XX 11-APR-2001; 2001WO-DK000264.
XX 17-APR-2000; 2000DK-00000651.
XX 29-MAY-2000; 2000DK-00000849.
XX (BIOI-) BIOIMAGE AS.
XX

PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66852.
XX
XX
PT Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX
PS Example 1; Page 156-160; 160pp; English.
XX
XX
CC The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
CC protein
XX
XX
SQ Sequence 893 AA;
Query Match 99.7%; Score 1269; DB 4; Length 893;
Best Local Similarity 99.6%; Pred. No. 6.9e-122;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFVSVEGEGDATYGLTLKFKICTTGKLPVPMPT 60
Db 655 MVSKEELFTGVVPILVELDGDVNGHKFVSVEGEGDATYGLTLKFKICTTGKLPVPMPT 714
Qy 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAFPEGYVQERTIFFKDDGNYKTRAEVKEFGDTL 120
Db 715 LVTLSYGVQCFSRYPDHMKQHDFFKSAFPEGYVQERTIFFKDDGNYKTRAEVKEFGDTL 774
Qy 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 775 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 834
Qy 181 DHYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLGFVTAAGITLGMDELYK 239
Db 835 DHYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLGFVTAAGITLGMDELYK 893
RESULT 4
ID AAG65782
XX AAG65782 standard; protein; 1132 AA.
XX
AC AAG65782;
XX
DT 07-JAN-2002 (first entry)
XX
DE Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
KW PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW fusion protein.
XX
OS Homo sapiens.
XX Aequorea victoria.
XX
XX WO200179526-A2.
XX
XX

PD 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-DK000264.
XX PF
XX PR
XX 17-APR-2000; 2000DK-00000651.
XX PR
XX 29-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
XX Praestegaard M;
XX
XX WPI; 2001-611727/70.
XX N-PSDB; AAI66853.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
XX compounds for treating CNS and inflammatory disease comprises identifying
XX compounds which remove PDE4 spots.
XX
XX Example 1; Page 162-167; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
XX protein
XX
XX SQ Sequence 1132 AA;
Query Match 99.7%; Score 1269; DB 4; Length 1132;
Best Local Similarity 99.6%; Pred. No. 9.9e-122;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFVSVEGEGDATYGLTLKFKICTTGKLPVPMPT 60
Db 894 MVSKEELFTGVVPILVELDGDVNGHKFVSVEGEGDATYGLTLKFKICTTGKLPVPMPT 953
Qy 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAFPEGYVQERTIFFKDDGNYKTRAEVKEFGDTL 120
Db 954 LVTLSYGVQCFSRYPDHMKQHDFFKSAFPEGYVQERTIFFKDDGNYKTRAEVKEFGDTL 1013
Qy 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 1014 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 1073
Qy 181 DHYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLGFVTAAGITLGMDELYK 239
Db 1074 DHYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLGFVTAAGITLGMDELYK 1132
RESULT 5
AAE17517
ID AAE17517 standard; protein; 239 AA.
XX
XX AAE17517;
XX
XX 22-APR-2002 (first entry)
XX
XX Enhanced F64L jellyfish green fluorescent protein mutant.
XX
XX

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
 KW cellular function; genetic reporter; mutant; Stoke's shift; mtein.
 OS Aequorea victoria.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 65
 FT /note= "wild type Phe substituted with Leu; This
 FT corresponds to position 64 in the wild type protein"
 XX
 PN WO200198338-A2.
 XX
 PD 27-DEC-2001.
 XX
 XX 18-JUN-2001; 2001WO-EP006848.
 XX
 PR 19-JUN-2000; 2000DK-00009953.
 PR 20-JUN-2000; 2000US-0212681P.
 PR 10-MAY-2001; 2001DK-00000739.
 PR 10-MAY-2001; 2001US-0290170P.
 XX
 XX (BIOI-) BIOIMAGE AS.
 PA
 XX Bjorn SP, Pagliaro L, Thastrup O;
 PI
 XX WPI; 2002-098224/13.
 DR N-PSDB; AAD28162.
 DR
 XX Novel fluorescent protein in in vitro assay for measuring protein kinase
 PT activity or dephosphorylation activity, or for measuring protein
 PT redistribution, has a green fluorescent protein with F64L and E222G
 PT mutation.
 XX
 PS Example 1; Page 35; 41pp; English.
 XX
 CC The invention relates to a fluorescent protein derived from green
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
 CC at F64L and E222G has a bigger compared to other GFP's making it very
 CC suitable for high throughput screening due to better resolution. The
 CC fluorescent protein is useful in invitro assays for measuring protein
 CC kinase activity or dephosphorylation activity, or for measuring protein
 CC redistribution. The fluorescent protein is useful in studying cellular
 CC functions in living cells; as protein tags in transgenic animals, living
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.
 CC The fluorescent protein is also useful as a cell or organelle integrity
 CC marker, a marker for changes in cell morphology, as transfection marker,
 CC and as a marker to be used in combination with fluorescence activated
 CC cell sorting (FACS). The novel proteins can also be used as reporters to
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
 CC protein is also useful as markers in transcriptional and translational
 CC fusions for performing transposon vector mutagenesis and as a reporter
 CC for bacterial detection. Transposons encoding the fluorescent protein are
 CC useful for screening promoters and for tagging plasmids and chromosomes.
 CC The fluorescent protein engineered into the genome of a phage is useful
 CC for designing diagnostic tool. The present sequence is enhanced F64L
 CC jellyfish green fluorescent protein (GFP) mutant
 XX
 SQ Sequence 239 AA;
 Query Match 99.1%; Score 1261; DB 5; Length 239;
 Best Local Similarity 99.2%; Pred. No. 6.5e-122;
 Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MYSKGELFTGVVPIVLVDGDNVGNHKFVSVSSEGGDGYGKLTLPKICTTGKLPVWPPT 60
 DB 1 MYSKGELFTGVVPIVLVDGDNVGNHKFVSVSSEGGDGYGKLTLPKICTTGKLPVWPPT 60
 QY 51 LVTLSGVQCFSRYPDHMKQHDFFKSAEMPEGVYVQERTIFFKDDGNKTRAEVKFEGDTL 120
 DB 51 LVTLSGVQCFSRYPDHMKQHDFFKSAEMPEGVYVQERTIFFKDDGNKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180

Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITTLGMDELYK 239
 Db 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITTLGMDELYK 239
 RESULT 6
 AAB222882
 ID AAB222882 standard; protein; 239 AA.
 XX
 AC AAB222882;
 XX
 DT 10-JAN-2001 (first entry)
 XX
 DE Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.
 XX
 KW Bioreactor protein; fusion protein; recognition site;
 KW cellular targeting sequence; cellular localisation; fluorescent protein;
 KW protease activity detection; toxin detection; cellular stress detection;
 KW drug discovery; cell based screening.
 XX
 OS Aequorea victoria.
 OS Synthetic.
 XX
 PN WO200050872-A2.
 XX
 PD 31-AUG-2000.
 XX
 XX 25-FEB-2000; 2000WO-US004794.
 PR 26-FEB-1999; 99US-0122152P.
 PR 08-MAR-1999; 99US-0123399P.
 PR 12-JUL-1999; 99US-00352171.
 XX
 XX (CELL-) CELLOMICS INC.
 PA
 XX Giuliano KA, Kapur R;
 PI
 XX WPI; 2000-594086/56.
 DR N-PSDB; AAA93373.
 XX
 PT Automated cell-based characterization of toxin by contacting cells
 PT containing luminescent reporter molecules with test substance and
 PT analyzing optically.
 XX
 PS Example 11; Fig 29A; 336pp; English.
 XX
 CC The invention relates to systems, methods and reagents for cell-based
 CC screening or detection of compounds which affect particular biological
 CC functions. The methods of the invention utilise fluorescent bioreactor
 CC molecules which, when acted on by a compound of interest, cause an
 CC alteration in the cellular distribution of at least the fluorescent
 CC moiety. In one embodiment, the biosensors comprise heat shock proteins
 CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent
 CC protein (GFP), or derivatives thereof). Such biosensors are located in
 CC the cytoplasm, but on stress activation translocate to the nucleus. In
 CC another embodiment bioreactor proteins can be used to detect protease
 CC activity. Such protease bioreactor fusion proteins comprise one or more
 CC fluorescent proteins; a recognition signal which is cleaved by the
 CC protease; and at least one cellular localisation signal. The latter two
 CC components may be components of a single protein which is acted upon by
 CC the protease, or may be from heterologous sources. Due to the
 CC localisation signal, the bioreactor protein is localised to a particular
 CC region of the cell. Once acted on by the protease of interest, the
 CC fluorescent protein is cleaved from the localisation sequence, and is
 CC free to migrate to other locations within the cell. The presence of a
 CC second localisation signal attached to the fluorescent protein enables
 CC the fluorescent protein to be directed to a different cellular
 CC compartment after cleavage of the protease recognition sequence. The
 CC change in distribution of the fluorescent protein can be detected using
 CC imaging methods with a high degree of spatial resolution. The methods and

CC biosensors of the invention can be used to investigate a wide range of
CC cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased quantity
CC and speed of data collection and faster evaluation of drug candidates.
CC Sequences AAB2881-B2885 represent fluorescent proteins which may be used
CC as components of biosensor fusion proteins of the invention
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVWPWT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 7
AAY54349
ID AAY54349 standard; protein; 239 AA.
XX AAY54349;
DT 06-APR-2000 (first entry)
XX Amino acid sequence of the mutant green fluorescent protein EGFP.
DE Fluorescent protein; green fluorescent protein; emission intensity;
KW fluorescence; pH detection; pH sensor; EGFP.
XX Synthetic.
OS Aequorea victoria.

XX Key Location/Qualifiers
FH Misc-difference 65 /note= "wild type Phe substituted with Leu"
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"
FT Misc-difference 232 /note= "wild type His substituted with Leu"
FT WO964592-A2.
XX
PN 16-DEC-1999.
XX
PD 08-JUN-1999; 99WO-US012850.
XX
PF 09-JUN-1998; 98US-00094359.
XX
PR 13-OCT-1998; 98US-00172063.
XX
XX (REGC) UNIV CALIFORNIA.
PA (UYOR-) UNIV OREGON STATE.
XX

PI Tsien RY, Llopis J, Wachter RM;
XX WPI; 2000-116540/10.
DR N-PSDB; AA245642.
XX New functional engineered green fluorescent proteins, used for measuring
PT the pH in biological samples and cells.
XX Disclosure; Page 9; 89pp; English.
XX The present sequence represents a functional engineered fluorescent
CC protein based on the Aequorea green fluorescent protein (GFP). The
CC emission intensity changes as pH varies between 5 and 10 of the present
CC protein are novel. The functional engineered fluorescent proteins show
CC reversible changes in fluorescence over physiological pH ranges. They can
CC be used for determining the pH of samples and cells. The polynucleotides
CC can also be used to produce transgenic animals. The fluorescent protein
CC pH sensors can be delivered to cells in the form of polynucleotides
CC encoding the protein sensor fused to a targeting signal. The targeting
CC signal directs the expression of the protein sensors to restricted cell
CC locations. This makes it possible to measure the pH of a precisely
CC defined cellular region or organelle
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVWPWT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 8
AAY79584
ID AAY79584 standard; peptide; 239 AA.
XX AAY79584;
XX 29-AUG-2000 (first entry)
XX EGFP signal domain.
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;
KW analysis; drug discovery.
XX Unidentified.
OS WO200026408-A2.
XX
PN 11-MAY-2000.
XX
PD 29-OCT-1999; 99WO-US025431.
XX
PF 30-OCT-1998; 98US-0106308P.
XX
PR 26-MAY-1999; 99US-0136078P.
XX
XX (CELL-) CELLOMICS INC.
XX
XX Guillian KA, Bright G, Olson K, Burroughs-Tencza S;

XX WPI; 2000-365644/31.
DR N-PSDB; AAA27573.
XX Recombinant nucleic acid encoding a protease biosensor useful for
PT fluorescence based cell and molecular biochemical assays for drug
PT discovery comprising three operably linked nucleic acid sequences.
XX
XX Claim 14; Fig 29A; 218pp; English.
XX
XX The present sequence is that of the EGFP signal domain, which can be
CC included in novel recombinant protease biosensors (PBs) of the invention.
CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)
CC comprising at least 1 detectable polypeptide signal such as the present
CC sequence; a second domain (see AAY79588-622) comprising at least 1
CC protease recognition site; and a third domain (see AAY79623-37)
CC comprising at least 1 reactant target sequence. A recombinant nucleic
CC acid (see AAA27627-43) encoding the PB, an expression vector, and a
CC genetically engineered host cell are also claimed. A claimed method for
CC identifying compounds that modify protease activity in a cell involves
CC contacting a host cell that possesses the recombinant PB with a test
CC compound, and determining the PB distribution in the host cell, where
CC changes in the distribution of the PB are correlated with modification of
CC protease activity by the test compound. Claimed kits for identifying
CC compounds that modify protease activity in a host cell include the
CC recombinant nucleic acid, or the recombinant PB, or the vector, or the
CC host cell. The PB is useful in high content screens to detect in vivo
CC activation of enzymatic activity, and to identify specific activity based
CC on cleavage of a known recognition motif
XX
XX Sequence 239 AA;
SQ

Query Match 98.8%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
QY 61 LVTVLVSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 131 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9
AAB50804
ID AAB50804 standard; protein; 239 AA.
XX
XX AAB50804;
XX
XX 14-MAR-2001 (first entry)
XX
XX Jellyfish GFP mutant EGFP.
XX
XX Aequorea victoria; jellyfish; fluorescent protein indicator;
KW green fluorescent protein; GFP; linker moiety; sensor;
KW calmodulin-binding domain; mutant; mutein.
XX
XX Aequorea victoria.
OS
XX WO200071565-A2.
XX
XX 30-NOV-2000.
XX

PF 17-MAY-2000; 2000WO-US013684.
XX
XX 21-MAY-1999; 99US-00316919.
PR 21-MAY-1999; 99US-00316920.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Tsien RV, Baird GA;
XX
XX WPI; 2001-032017/04.
DR N-PSDB; AAC90488.
XX
XX Novel fluorescent proteins comprising a sensor protein inserted into
PT them, useful for measuring the response of a sensor biological, chemical,
PT electrical or physiological parameter in vivo or in vitro.
XX
XX Disclosure; Page 24; 94pp; English.
PS
XX The present sequence is a fluorescent protein used in the construction of
CC a fluorescent protein indicator. The indicator comprises a sensor
CC polypeptide that is responsive to a chemical, biological, electrical or
CC physiological parameter, and a fluorescence protein functional group. The
CC sensor polypeptide is operatively inserted into the fluorescent moiety.
CC The fluorescent indicator is useful for detecting the presence of a
CC response inducing member in a sample. The method involves contacting the
CC sample with the indicator and detecting a change in fluorescence, in
CC which a change is indicative of the effect of the parameter on the sensor
CC polypeptide. The novel fluorescent proteins are advantageous due to their
CC reduced size as compared to the FRET (fluorescence resonance energy
CC transfer)-based sensors
XX
XX Sequence 239 AA;
SQ

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
QY 61 LVTVLVSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 10
AAB85900
ID AAB85900 standard; protein; 239 AA.
XX
XX AAB85900;
XX
XX 30-NOV-2001 (first entry)
XX
XX A. victoria green fluorescent protein (GFP) and linker sequence.
XX
XX Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;
KW green fluorescent protein.
XX
XX Synthetic.
OS
XX Aequorea victoria.
XX
XX WO200168706-A1.
XX


```

PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-US008071.
XX
PR 15-MAR-2000; 2000US-0189698P.
XX
PA (MERI ) MERCK & CO INC.
XX
XX Marsh DJ;
XX
XX WPI; 2001-565791/63.
DR N-PSDB; AAH47304.
XX
XX Fusion proteins comprising melanin concentrating hormone receptor
PT peptides and fluorescent proteins, useful for identifying appetite
PT stimulants.
XX
XX Claim 2; Page 14; 71pp; English.
XX
XX The invention provides melanin concentrating hormone (MCH) receptor
CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
CC MCHR polypeptide regions from different species. The MCHR fusion protein
CC comprise MCHR polypeptide region and a fluorescent polypeptide region
CC joined directly, or via a linker, to the carboxy side of the MCHR
CC polypeptide region. The MCHR fusion proteins can be expressed by standard
CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
CC regulation of MCH activity stimulates food intake. The present sequence
CC represents a A. victoria green fluorescent protein (GFP) and a linker
CC sequence
XX
XX Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTILKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTILKFICTTGKLPVWPWT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDPFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 11
AAB31171
XX ID AAB31171 standard; protein; 239 AA.
XX AC AAB31171;
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a green fluorescent protein (GFP).
XX
XX Growth rate; death rate; reporter gene; luminescent protein;
KW fluorescent product; luciferase; green fluorescent protein; GFP.
XX
XX Aequorea victoria.
XX
XX WO200075367-A1.
XX
XX 14-DEC-2000.
XX
XX 07-JUN-2000; 2000WO-FI000507.

XX
XX 07-JUN-1999; 99FI-00001296.
XX
XX (LILI/) LILIUS E.
PA (VIRT/) VIRTA M.
XX
XX Lilius E, Virta M;
XX
XX WPI; 2001-061737/07.
DR N-PSDB; AAC86954.
XX
XX Assessing growth and death rates of a micro-organism in a desired
PT environment, by introducing 2 reporter genes encoding luminescent and
PT fluorescent products and detecting luminescent fluorescence.
XX
XX Disclosure; Page 27; 32pp; English.
XX
XX The specification describes a method for assessing the growth rate and
CC death rate of a micro-organism within a predetermined time period in a
CC desired environment. The method comprises introducing at least two
CC reporter genes encoding luminescent and/or fluorescent products into the
CC micro-organisms, incubating the micro-organism within the desired
CC environment, and detecting luminescence and/or fluorescence after a
CC predetermined time period. Use of two different markers within a micro-
CC organism enables the differentiation between growth and death rates. The
CC method is used to assess the growth rate and death rate of a micro-
CC organism within a predetermined time period in a desired environment. The
CC present sequence represents a green fluorescent protein (GFP), and is
CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
CC and is used in the method of the invention
XX
XX Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTILKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTILKFICTTGKLPVWPWT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDPFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 12
AAG66198
XX ID AAG66198 standard; protein; 239 AA.
XX AC AAG66198;
XX
XX 17-JUN-2002 (first entry)
XX
XX A. victoria green fluorescent protein (EGFP).
XX
XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
KW green fluorescent protein; EGFP.
XX
XX Aequorea victoria.
XX
XX JP2002045189-A.
XX
XX 12-FEB-2002.
XX

```


XX Key Location/Qualifiers
 FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"
 FT
 FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"
 FT FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"
 FT
 XX EP1178109-AL.
 XX
 PD 06-FEB-2002.
 XX
 XX 03-AUG-2001; 2001EP-00306650.
 XX
 PR 04-AUG-2000; 2000JP-00237166.
 XX
 PA (RIKE) RIKEN KK.
 XX
 PI Miyawaki A, Sawano A;
 XX
 DR WPI; 2002-208112/27.
 DR N-PSDB; AAD27910.
 XX
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimer.
 PT
 PT
 PS Example 1; Page 13-14; 31pp; English.
 XX
 XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimer. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments, and a double-stranded of digestion as megaprimer, the megaprimer are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria
 XX
 SQ Sequence 239 AA;
 Query Match 98.8%; Score 1258; DB 5; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.3e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MWSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPVWPWT 60
 DB 1 MWSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPVWPWT 60
 QY 61 LVTLVSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGDIDFKEDGNILGHLEYNYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGDIDFKEDGNILGHLEYNYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
 RESULT 15
 AAE34958
 ID AAE34958 standard; protein; 239 AA.
 XX
 AC AAE34958;
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Aequorea victoria enhanced green fluorescent protein (EGFP).
 XX
 KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;
 KW Kinase; enhanced green fluorescent protein; EGFP.
 XX
 OS Aequorea victoria.
 XX
 PN WO200295058-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 24-MAY-2002; 2002WO-USO16955.
 XX
 PR 24-MAY-2001; 2001US-00865291.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Tsien RY, Ting AV, Zhang J;
 XX
 DR WPI; 2003-148474/14.
 DR N-PSDB; AAD53428.
 XX
 PT Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.
 PT
 XX Disclosure; Col 56-57; 38pp; English.
 XX
 CC The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention
 XX
 SQ Sequence 239 AA;
 Query Match 98.8%; Score 1258; DB 6; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.3e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MWSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPVWPWT 60
 DB 1 MWSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPVWPWT 60
 QY 61 LVTLVSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGDIDFKEDGNILGHLEYNYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGDIDFKEDGNILGHLEYNYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

Db 181 DHYQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:09
Job time : 47.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-64v
Perfect score: 1273
Sequence: 1 MVSKEELFTGVVPIVLVELD.....VLLGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCFUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfilesi.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1258	98.8	239	3 US-09-172-063-3	Sequence 3, Appli
2	1258	98.8	239	4 US-09-513-783A-46	Sequence 46, Appli
3	1258	98.8	239	4 US-09-316-919-4	Sequence 4, Appli
4	1258	98.8	239	4 US-09-602-641-3	Sequence 3, Appli
5	1258	98.8	239	4 US-09-920-922-2	Sequence 2, Appli
6	1258	98.8	281	3 US-09-062-102-1	Sequence 1, Appli
7	1258	98.8	281	4 US-09-364-946-1	Sequence 1, Appli
8	1258	98.8	294	4 US-09-513-783A-2	Sequence 2, Appli
9	1258	98.8	323	3 US-09-172-063-21	Sequence 21, Appli
10	1258	98.8	323	4 US-09-602-641-21	Sequence 21, Appli
11	1258	98.8	364	3 US-09-085-305-6	Sequence 6, Appli
12	1258	98.8	379	4 US-09-417-197-129	Sequence 129, App
13	1258	98.8	434	4 US-09-800-170-48	Sequence 48, Appli
14	1258	98.8	442	4 US-09-417-197-127	Sequence 127, App
15	1258	98.8	459	4 US-09-513-783A-170	Sequence 170, App
16	1258	98.8	544	4 US-09-417-197-113	Sequence 113, App
17	1258	98.8	544	4 US-09-417-197-115	Sequence 115, App
18	1258	98.8	604	4 US-09-417-197-59	Sequence 59, Appli
19	1258	98.8	605	4 US-09-417-197-41	Sequence 41, Appli
20	1258	98.8	606	4 US-09-417-197-65	Sequence 65, Appli
21	1258	98.8	607	4 US-09-417-197-47	Sequence 47, Appli
22	1258	98.8	630	4 US-09-417-197-63	Sequence 63, Appli
23	1258	98.8	631	4 US-09-417-197-39	Sequence 39, Appli
24	1258	98.8	633	4 US-09-417-197-45	Sequence 45, Appli
25	1258	98.8	635	4 US-09-417-197-125	Sequence 125, App
26	1258	98.8	642	2 US-08-818-253-2	Sequence 2, Appli
27	1258	98.8	642	2 US-08-818-253-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-172-063-3
; Sequence 3, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-172-063-3

Query Match 98.8%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy	1	MVSKGEELFTGVVPIVLVELDGDVNGHKFSVSGEGDATYVGLTKLKFICTTGKLPVWPWT	60
Db	1	MVSKGEELFTGVVPIVLVELDGDVNGHKFSVSGEGDATYVGLTKLKFICTTGKLPVWPWT	60
Qy	61	LVTVLTVGVQCFSRYPDHNKQHDFFKSAKMPGVQVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Db	61	LVTTLTVGVQCFSRYPDHNKQHDFFKSAKMPGVQVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Qy	121	VNRIELKIDDFKEDGNILGHKLEYNVSNHNYIMADKQKNGIKVNFKIRHNIEDGSVOLA	180
Db	121	VNRIELKIDDFKEDGNILGHKLEYNVSNHNYIMADKQKNGIKVNFKIRHNIEDGSVOLA	180
Qy	181	DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK	239

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTALKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTALKFICTTGKLPVPWPT 60
QY 51 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 51 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHNLSQALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQONTPIGDGPVLLPDNHNLSQALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTALKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTALKFICTTGKLPVPWPT 60
QY 51 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 51 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHNLSQALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQONTPIGDGPVLLPDNHNLSQALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wichter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTALKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTALKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNLSQALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNLSQALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-920-922-2

Query Match      98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match      98.8%; Score 1258; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.7e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match      98.8%; Score 1258; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.7e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match      98.8%; Score 1258; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 1.8e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
```

Db 1 MVSKEBELFTGVVPIVLVLDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60
QY 61 LVTVLGYVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9

US-09-172-063-21
; Sequence 21, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-172-063-21

Query Match 98.8%; Score 1258; DB 3; Length 323;
Best Local Similarity 98.7%; Pred. No. 2.1e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEBELFTGVVPIVLVLDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60
Db 85 MVSKEBELFTGVVPIVLVLDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 144
QY 61 LVTVLGYVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21
; Sequence 21, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-602-641-21

Query Match 98.8%; Score 1258; DB 4; Length 323;
Best Local Similarity 98.7%; Pred. No. 2.1e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEBELFTGVVPIVLVLDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60
Db 85 MVSKEBELFTGVVPIVLVLDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 144
QY 61 LVTVLGYVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6
; Sequence 6, Application US/09085305
; Patent No. 6191269
; GENERAL INFORMATION:
; APPLICANT: Pollock, Allan
; APPLICANT: Lovett, David H.
; APPLICANT: Turck, Johanna
; TITLE OF INVENTION: Selective Induction of Apoptosis in
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESS: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,305
; FILING DATE: 29-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Francis, Carol L
REGISTRATION NUMBER: 36,513
REFERENCE/DOCKET NUMBER: 6510/102US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-085-305-6

Query Match 98.8%; Score 1258; DB 3; Length 364;
Best Local Similarity 98.7%; Pred. No. 2.5e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTCLKFICTTGTGKLPVPWPT 60
DB 126 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTCLKFICTTGTGKLPVPWPT 185

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 186 LVTTLTYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 246 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 305

QY 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 306 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 364

RESULT 12
US-09-417-197-129
Sequence 129, Application US/09417197
Patent No. 6518021
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE OF INVENTION: On A Cellular Response
FILE REFERENCE: 3759-0110P
CURRENT APPLICATION NUMBER: US/09/417,197
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 129
LENGTH: 379
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: actin-binding-domain-EGFP fusion
US-09-417-197-129

Query Match 98.8%; Score 1258; DB 4; Length 379;
Best Local Similarity 98.7%; Pred. No. 2.7e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTCLKFICTTGTGKLPVPWPT 60
DB 141 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTCLKFICTTGTGKLPVPWPT 200

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 201 LVTTLTYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 261 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 320

QY 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 321 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 379

RESULT 13
US-09-800-170-48
Sequence 48, Application US/09800170
Patent No. 6481667
GENERAL INFORMATION:
APPLICANT: Kinsella, Todd
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
FILE REFERENCE: A-68614-1/DJB/RMS/RMK
CURRENT APPLICATION NUMBER: US/09/800,170
CURRENT FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: US 60/187,130
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 90
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48
LENGTH: 434
TYPE: PRT
ORGANISM: Synechocystis PCC6803
US-09-800-170-48

Query Match 98.8%; Score 1258; DB 4; Length 434;
Best Local Similarity 98.7%; Pred. No. 3.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTCLKFICTTGTGKLPVPWPT 60
DB 196 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTCLKFICTTGTGKLPVPWPT 255

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 256 LVTTLTYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 316 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 375

QY 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 376 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 434

RESULT 14
US-09-417-197-127
Sequence 127, Application US/09417197
Patent No. 6518021
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE OF INVENTION: On A Cellular Response
FILE REFERENCE: 3759-0110P
CURRENT APPLICATION NUMBER: US/09/417,197
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 127
LENGTH: 442
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EGFP-RhoA fusion
US-09-417-197-127

Query Match 98.8%; Score 1258; DB 4; Length 442;
Best Local Similarity 98.7%; Pred. No. 3.4e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTCLKFICTTGTGKLPVPWPT 60

Db	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT	60
Qy	61	LVTVLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNGYKTRAEVKFEGDTL	120
Db	61	LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNGYKTRAEVKFEGDTL	120
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK	239

RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170

Query Match	98.8%	Score 1258;	DB 4;	Length 459;
Best Local Similarity	98.7%	Pred. No. 3.5e-126;		
Matches	236;	Conservative	1;	Mismatches 2; Indels 0; Caps 0;
Qy	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT	60	
Db	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT	60	
Qy	61	LVTVLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNGYKTRAEVKFEGDTL	120	
Db	61	LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNGYKTRAEVKFEGDTL	120	
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180	
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180	
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK	239	
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK	239	

Search completed: June 21, 2004, 16:04:01
Job time : 12.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds
(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-64v

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPIVLVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1269	99.7	239	9	US-09-887-784-4
2	1269	99.7	239	12	US-10-296-953-4
3	1269	99.7	363	14	US-10-270-223-6
4	1269	99.7	893	14	US-10-257-909A-30
5	1269	99.7	1132	14	US-10-257-909A-32
6	1261	99.1	239	9	US-09-887-784-2
7	1261	99.1	239	12	US-10-296-953-2
8	1258	98.8	239	9	US-09-920-922-2
9	1258	98.8	239	9	US-09-999-745-4
10	1258	98.8	239	10	US-09-866-538-4
11	1258	98.8	239	10	US-09-797-496B-2
12	1258	98.8	239	10	US-09-794-308-4
13	1258	98.8	239	10	US-09-865-291-4
14	1258	98.8	239	12	US-10-457-982-3
15	1258	98.8	239	14	US-10-121-258-13

16	1258	98.8	239	14	US-10-221-461-7	Sequence 7, Appli
17	1258	98.8	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1258	98.8	239	14	US-10-177-390-2	Sequence 2, Appli
19	1258	98.8	239	14	US-10-338-411-3	Sequence 3, Appli
20	1258	98.8	239	15	US-10-370-570-4	Sequence 4, Appli
21	1258	98.8	239	15	US-10-389-640-3	Sequence 3, Appli
22	1258	98.8	259	14	US-10-314-861-11	Sequence 11, Appli
23	1258	98.8	281	12	US-09-931-232-1	Sequence 1, Appli
24	1258	98.8	288	14	US-10-314-861-37	Sequence 37, Appli
25	1258	98.8	293	14	US-10-314-861-35	Sequence 35, Appli
26	1258	98.8	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1258	98.8	295	14	US-10-314-861-39	Sequence 39, Appli
28	1258	98.8	299	14	US-10-314-861-33	Sequence 33, Appli
29	1258	98.8	305	14	US-10-314-861-31	Sequence 31, Appli
30	1258	98.8	308	14	US-10-033-717-35	Sequence 35, Appli
31	1258	98.8	311	14	US-10-314-861-29	Sequence 29, Appli
32	1258	98.8	320	14	US-10-338-411-11	Sequence 11, Appli
33	1258	98.8	320	15	US-10-389-640-11	Sequence 11, Appli
34	1258	98.8	323	12	US-10-457-982-21	Sequence 21, Appli
35	1258	98.8	323	14	US-10-338-411-7	Sequence 7, Appli
36	1258	98.8	323	14	US-10-338-411-13	Sequence 13, Appli
37	1258	98.8	323	15	US-10-389-640-7	Sequence 7, Appli
38	1258	98.8	323	15	US-10-389-640-13	Sequence 13, Appli
39	1258	98.8	324	14	US-10-314-861-16	Sequence 16, Appli
40	1258	98.8	345	14	US-10-338-411-5	Sequence 5, Appli
41	1258	98.8	345	15	US-10-389-640-5	Sequence 5, Appli
42	1258	98.8	346	14	US-10-338-411-9	Sequence 9, Appli
43	1258	98.8	346	15	US-10-389-640-9	Sequence 9, Appli
44	1258	98.8	359	14	US-10-033-717-33	Sequence 33, Appli
45	1258	98.8	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: EUBORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequoria Victoria
US-09-887-784-4

Query Match 99.7%; Score 1269; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 7.3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVWPWT	60
Db	1	MVSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVWPWT	60
Qy	61	LVTVLISYGVQCFSRYPDHMKQHDFFPKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Db	61	LVTTLISYGVQCFSRYPDHMKQHDFFPKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Qy	121	VNRIELKIDFDKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGVSQLA	180
Db	121	VNRIELKIDFDKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGVSQLA	180
Qy	181	DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGVFTVTAAGITLGMDELYK	239
Db	181	DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGVFTVTAAGITLGMDELYK	239

```
RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match          99.7%; Score 1269; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 7.3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
QY 51 LVTVLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 51 LVTVLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITIGMDLYK 239
DB 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITIGMDLYK 239

RESULT 3
US-10-270-223-6
; Sequence 5, Application US/10270223
; Publication No. US2003014363A1
; GENERAL INFORMATION:
; APPLICANT: Bioimage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM THE SAME
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match          99.7%; Score 1269; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 1.3e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
QY 51 LVTVLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 51 LVTVLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITIGMDLYK 239
DB 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITIGMDLYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.7%; Score 1269; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 4.8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 714
QY 61 LVTVLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 715 LVTVLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITIGMDLYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITIGMDLYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.7%; Score 1269; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 6 8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
DB 894 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 953

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFDKEDGNILGHKLEYNHNYSYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US2002017189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          99.1%; Score 1261; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 5e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
```

```
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.1%; Score 1261; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 5e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNYSYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.8%; Score 1258; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60
```


APPLICANT: TSJEN, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match 98.8%; Score 1258; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
Sequence 4, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSJEN, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REGEN1550
CURRENT APPLICATION NUMBER: US/09/865,291
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match 98.8%; Score 1258; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US20030212265A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Atsushi
APPLICANT: Llopis, Juan
APPLICANT: Wächter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR FILING DATE: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 98.8%; Score 1258; DB 12; Length 239;
Best Local Similarity 98.7%; Pred. No. 1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 15
US-10-121-258-13
Sequence 13, Application US/10121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP1
CURRENT APPLICATION NUMBER: US/10/121,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.8%; Score 1258; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDGYGKLTTLKFICTTGKLPVPWPT 60
Db      1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDGYGKLTTLKFICTTGKLPVPWPT 60

QY      61 LVTLSVGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db      61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMYLLGFVTAAGITLGMDELYK 239
Db      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMYLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:27
Job time : 35.7778 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds
(without alignments)
2224.817 Million cell updates/sec

Title: US-09-887-784-64V

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPILVELD.....VLLGFVTRAGITLGMDELTK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1236	97.1	238	1 JQ1514	green-fluorescent
2	1105	8.2	785	2 H72228	hypothetical prote
3	93.5	7.3	861	2 H64102	leucine-tRNA synth
4	91.5	7.2	887	2 E82590	leucyl-tRNA synth
5	89.5	7.0	860	2 AC0582	DNA topoisomerase
6	88.5	7.0	655	2 D83917	iron-sulfur cofact
7	87.5	6.9	370	2 E70390	hypothetical prote
8	87.5	6.9	2573	2 D71614	DNA-directed DNA p
9	86.5	6.8	788	1 JCVLHH	protective surface
10	86.5	6.8	797	2 JDC4078	protective surface
11	86.5	6.8	808	2 F64102	tRNA (uracil-5)-m
12	86	6.8	357	2 G81355	photosystem II chl
13	86	6.8	461	2 T06936	inter-alpha-trypsi
14	85.5	6.7	889	2 JC5576	cellulase (EC 3.2.
15	85.5	6.7	941	2 S29043	hypothetical prote
16	85	6.7	281	2 AD2052	hypothetical prote
17	85	6.7	336	2 C64468	synaptogamin o-p65
18	85	6.7	439	2 JH0414	DNA-binding protei
19	85	6.7	632	2 T06586	oligodeoxyphosphat
20	84.5	6.6	613	2 A99552	DNA-directed DNA p
21	84.5	6.6	836	1 JCVLDD	thioredoxin reduct
22	84	6.6	285	2 C97279	dihydrolipamide d
23	84	6.6	578	1 I40794	hypothetical prote
24	84	6.6	682	2 D90946	hypothetical prote
25	84	6.6	682	2 H85794	hypothetical prote
26	84	6.6	682	2 A41798	carboxy-terminal p
27	83.5	6.6	425	2 C97354	IgA Fc receptor pr
28	83.5	6.6	1134	2 A60234	IgA Fc receptor pr
29	83.5	6.6	1164	1 FC5OAG	

ALIGNMENTS

RESULT 1

JQ1514

Green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)

C:Species: Aequorea victoria

C>Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001

C:Accession: JQ1514; PQ0335; S48693; S51330; S51331

R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.

Gene 111, 229-233, 1992

A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.

A:Reference number: JQ1514; MUID:92175527; PMID:1347277

A:Accession: JQ0692

A:Molecule type: DNA

A:Residues: 1-107, 'S', 109-238 <PRAL>

A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663

A:Accession: JQ1514

A:Molecule type: mRNA

A:Residues: 1-99, 'P', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>

A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661

A:Accession: PQ0335

A:Molecule type: protein

A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>

R:Inouye, S.; Tsuji, F.I.

FEBS Lett. 351, 211-214, 1994

A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Accession: S48693

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>

A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384

R:Watkins, J.N.; Campbell, A.K.

submitted to the EMBL data Library, January 1995

A:Reference number: S51330

A:Accession: S51330

A:Molecule type: mRNA

A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>

A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009

A:Experimental source: clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 209-238 <PRA5>

A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011

A:Experimental source: clone gfp2

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-91

A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Nat. Biotechnol. 14, 1246-1251, 1996

A:Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

leucine-tRNA ligase
leucine tRNA synth
leucine tRNA synth
DNA-directed DNA p
imidazoleglycerol-
hypothetical prote
proprotein convert
hypothetical prote
alpha-amylase homo
lipoxigenase (EC 1
S-layer protein pr
water-stress-induc
neuraminidase, pro
synergohemotropi
ferrichrome-bindin
hypothetical prote

30 82.5 6.5 860 1 SYECL
31 82.5 6.5 860 2 H90713
32 82.5 6.5 860 2 D85564
33 82.5 6.5 2222 1 A36028
34 82 6.4 353 2 E84941
35 82 6.4 471 2 T27856
36 82 6.4 752 1 KXRTCL
37 82 6.4 764 2 T25012
38 82 6.4 774 2 T39539
39 82 6.4 865 2 T11852
40 82 6.4 874 2 JC4930
41 81.5 6.4 263 2 S53488
42 81.5 6.4 740 2 G95153
43 81 6.4 310 2 S68225
44 81 6.4 355 2 H97144
45 81 6.4 529 2 B86815

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF2176
C;Superfamily: leucine-tRNA ligase

Query Match 7.2%; Score 91.5; DB 2; Length 887;
Best Local Similarity 22.2%; Pred. No. 14;
Matches 44; Conservative 30; Mismatches 67; Indels 57; Gaps 10;

QY 50 TTGKLPVPMPTLVLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNY-- 107
Db TNEQLPV-WVANPVMAYGTGAVMAVEGHQDQDEF--ANKYGLPIRQVIALKEPKNQDE 385

QY 108 -----KTRAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVI 153
Db STWEPDWRDWAYADKTR---EPE---LINSAEFDGLDYQDAFEVLAERFE----- 429

QY 154 MADKQKNG-IKVNFKIRHNIEDGSVOLADHYQQTPI-----GDGPVLLPDN 199
Db 430 ---RQGRQRVNYRLR-----DWGVSQRQYWGCPPIVIYCTCGAVVPEDQLPVILPEN 482

QY 200 -HYLSTQSALSADPNRKR 216
Db 483 VAFSGTGSPIKTDPEWRK 500

RESULT 5
AC0582
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S
A;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called salmonella typhi
C;Date: 03-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0582
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0582
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-860 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176
C;Genetics:
A;Gene: STY0699
C;Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 860;
Best Local Similarity 23.3%; Pred. No. 20;
Matches 42; Conservative 22; Mismatches 77; Indels 39; Gaps 7;

QY 50 TTGKLPVPMPTLVLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKT 109
Db TGEIIPV-WVANPVMLEYGTGAVMAVPGH-QDQD-YEFASKYGLTIKPVILAADGSEPD 370

QY 110 RAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIR 169
Db SEQALTEKGVLFNSGFDGLAFAAFNATADKL-----REKVGGERKVNRYLR 418

QY 170 H-----NIEDGSVOLADHYQQTPTIGDGPVLLPDNHYL-STOSALSADP 212
Db DWGVSQRQYWGAPIPVMTLEDGT-----LPTPDQLPVILPDVMDGITSPIKADP 471

RESULT 6
DB3917
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: DB3917
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: AB3650; MUID:20512582; PMID:11058132
C;Accession: DB3917
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <STO>
A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05859.1; GSPDB:GN00
C;Experimental source: strain C-125
C;Genetics:
A;Gene: BH2140
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 7.0%; Score 88.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK---FSVSGEGEGDAT---YGKLTLPFI-----CTTGKLPVWP 59
Db 63 NVTIHKDQSVSRDEGRGMPGTGMHKLKGTPEVILTVLHAGGKFGGGYATSGLHCVGA 122

QY 60 TLVTVLSYGVCFSRYPDHMKQHDFFKSAPEGYVQER-----TIFKDDG----- 105
Db 123 SVNALSEWLIVEIKRDGWYEQRFENGKPKSTLEKKGKTRGTGTTIHKPDPTVFSST 182

QY 106 --NYKTRAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVIYIMADK----- 157
Db 183 NFNVETLSRLREAAFLKGLKIELVDLRDQDTEKVFH-YEDGIGKAFVEYLNEDEKTLHPV 241

QY 158 -----QKNGIKVNFKIRHNIEDGSVOLADHYQQTPTIGDGPVLLPDNHYLSTOSALSADP 212
Db 242 VFNESNGIEIEFAFQFN--DGYTENVLSFVNVTQDG-----GTHELCAKTMTRAV 294

QY 213 NE 214
Db 295 NE 296

RESULT 7
E70390
iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus
N;Contains: L-cysteine sulfotransferase (EC 2.8.1.-)
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999
C;Accession: E70390
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70390
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-370 <AQF>
A;Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065
C;Experimental source: strain VFS
C;Genetics:
A;Gene: nifs1
C;Superfamily: nitrogen fixation protein nifs
C;Keywords: phosphoprotein; pyridoxal phosphate; sulfotransferase
F;195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted
F;318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match 6.9%; Score 87.5; DB 2; Length 370;
Best Local Similarity 25.4%; Pred. No. 9.8;
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY 4 KGEELFTGVV-----PFLVELD-----GDVNGHKP-SVSGEG-----EGDATYKGLTLPFICT 50
Db 164 KGVPLLTDVAQIKGPIELKLNISATFSCHKFAIKGSGFLYISDEANYEPLIVGGQE 223

```
QY 51 TGKLP-----VPMPTLVLSYGVQCFSRYPDHMKQ-HDFEKSAMPEGYVQERTIFPKOD 104
Db 224 NGKRGSTENVVGLSLAKALEIIVNSFRYQEQKKURDLFENLLLEA-LPDAQIVGKDA 282
QY 105 GNYKTRAEV---KPEGDTLVNRIELKIGIDFKEDGNILGHKLEYNYNHNVIMADKOKNG 161
Db 283 ERSPSISVIMPRFFGAEIVNKLSEKGIYCGTSGACLSGEYENKMKLMKMGFSQEKALRM 342
QY 162 IKNVFKIRHNIED 174
Db 343 VRFSFGLNKEEE 355

RESULT 8
D71614
hypothetical protein PFB0460c - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C/Accession: D71614
R/Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Percec, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A/Reference number: A71600; MUID:99021743; PMID:9804551
A/Accession: D71614
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-2573 <GAR>
A/Cross-references: GB:AE001396; GB:AE001362; NID:g3845188; PIDN:AACT71881.1; PID:g384519
A/Experimental source: clone 3D7
C/Genetics:
A/Gene: PFB0460c

Query Match 6.8%; Score 87.5; DB 2; Length 2573;
Best Local Similarity 26.2%; Pred. No. 1.2e+02;
Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VQERTIFFKD--DNGYKTRAEVKEFGDTLVNRIELKIGIDFKEDGNILGHKLEYN--YNH 149
Db 126 LKKTETILCKDIKSGSNDPMDIEISLFDKDDWVDDKELK--DFEKSLLKIKKKEVNFYNNK 183
QY 150 NVYIMADKQKNGIKVNFKIRHNIEDGSGVQLADHYQNTPIGDGPVLLPNHLYLSTOSALS 209
Db 134 NLHIKENKKDEKKNKHNNNDNNM-----IYYKNI---DKTHYLDNNVVHILNDIN 236
QY 210 KDPNEKRDM 219
Db 237 TYLKRERDYM 246

RESULT 9
JDVLUHH
DNA-directed DNA polymerase (EC 2.7.7.7) - heron hepatitis virus
C/Species: heron hepatitis virus, HHV9
A/Note: host: Ardea cinerea (Gray heron)
C/Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 11-Jun-1999
C/Accession: A30082
R/Sprenger, R.; Kaleta, E.F.; Will, H.
J. Virol. 62, 3832-3839, 1988
A/Title: Isolation and characterization of a hepatitis B virus endemic in herons.
A/Reference number: A93037; MUID:88333160; PMID:3418788
C/Accession: A30082
A/Molecule type: DNA
A/Residues: 1-788 <SPR>
A/Cross-references: GB:M22056; NID:g325452; PIDN:AAA45738.1; PID:g325454
C/Superfamily: hepatitis virus DNA-directed DNA polymerase
C/Keywords: DNA biosynthesis; nucleotidyltransferase

Query Match 6.8%; Score 86.5; DB 1; Length 788;
Best Local Similarity 19.7%; Pred. No. 32;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;
```

```
QY 58 WPTLVTLVSYGVQCFSRYPDHMKQ-----DFFKSAMPEGYVQERT-----IFFKDDGNYK 108
Db 139 WPKSISTPLVHSGVKPKYFEFQNHESLVNDYLNKLFEAGILYKRVSKHLVTFK--GPYP 196
QY 109 T-----RAEVKEGDTLVNRIELKIGIDFKEDGNILGHKLEYNYNHNVIM 154
Db 197 TWEQKHLVPQOHGAYSKINDROESRRRIITATSSRKNDSSRI-----PCAHN--- 245
QY 155 ADKQKNGIKVNFKIRHNIEDGSGVQLADHYQ-----QNTPIGDGPVLL--PNHYL 202
Db 246 -----NGRKISY---HSTRDGGHRLSGRTSDPTSRGALAGGSDTPIGPGSTAHPSTHHV 297
QY 203 STQ-----SALSKDPNEKR 216
Db 298 DRRRROKGGVQLQAISRFESETR 320

RESULT 10
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C/Species: Haemophilus influenzae
A/Variety: type b
C/Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C/Accession: JC4078
R/Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A/Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl
A/Reference number: JC4078; MUID:95255676; PMID:7737523
A/Accession: JC4078
A/Molecule type: DNA
A/Residues: 1-797 <PLA>
A/Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448
A/Experimental source: type b
C/Superfamily: protective surface antigen D-15
C/Keywords: surface antigen
F/1-19/Domain: signal sequence #status predicted <SIG>
F/20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 6.8%; Score 86.5; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. No. 32;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY 65 LSGVQCFSRYPDHMKQHDF-----PKSAMPEGYVQE-----RTI 99
Db 427 IGVGESGISYQASVYKQDNFLGTGRAVSTAGTKNDYGTSVNLGYTEPYFTKDGVSIGNV 486
QY 100 FFKDDGNYKTRAEVKEFGDTLVNRIELKIGIDFKEDGNI---LGH-----KLEYNYS 148
Db 487 FFENYDNSKSDTSSNYKRTTYGNSVTL-GFPVNNNSYVVGILGHTYKNISNFALEYN--- 542
QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGSGVQLADHYQ-----NTPIGDGPVLL 196
Db 543 RNLYIQSMKFKNGIKTN-----DFDFSFGWYNLSLNRGYFPTKGVKASLG-GRVTI 593
QY 197 P--DNHLYLSTOSALS KDPNEKRDMVLLGCFVTAAGITLG 233
Db 594 PGSDNKYKLSADVQGFPLDRDLHLVWVSASAGYANG 632

RESULT 11
F64102
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C/Accession: F64102
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.L.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: F64102
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.4444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-64V

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPILVELD.....VLGLFVTAAGITLGMDELK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1242	97.6	238	1	GFP_AEQVI
2	93.5	7.3	861	1	SYL_HAEIN
3	91.5	7.2	879	1	SYL_XYLFA
4	89.5	7.0	860	1	SYL_XYLFA
5	89.5	7.0	860	1	SYL_SALTI
6	88.5	7.0	879	1	SYL_SALTY
7	86.5	6.8	533	1	CP51_CANGA
8	86.5	6.8	788	1	DPOL_HPEHE
9	86.5	6.8	795	1	D152_HAEIN
10	86.5	6.8	797	1	D151_HAEIN
11	86.5	6.8	859	1	SYL_SHEON
12	86	6.8	357	1	TRMA_CAWJE
13	86	6.8	461	1	PSBC_CVAPA
14	85.5	6.7	538	1	GRBE_RAT
15	85.5	6.7	793	1	ITH3_MESAU
16	85.5	6.7	886	1	ITH3_MESAU
17	85.5	6.7	941	1	GUN_EACS6
18	85.5	6.7	1603	1	VID4_CABEL
19	85	6.7	336	1	YD48_METJA
20	85	6.7	439	1	SY62_DISOM
21	84.5	6.6	501	1	AMPA_WIGBR
22	84.5	6.6	613	1	PEPF_WICPU
23	84	6.6	504	1	YC03_KLEPN
24	84	6.6	682	1	PRC_ECOLI
25	83.5	6.6	1164	1	BAG_STRAG
26	82.5	6.5	860	1	SYL_ECO57
27	82.5	6.5	860	1	SYL_ECOL6
28	82.5	6.5	860	1	SYL_ECOLI
29	82.5	6.5	2222	1	DPOE_YEAST
30	82	6.4	353	1	HIS7_BUCAI
31	82	6.4	689	1	AC2L_HUMAN
32	82	6.4	752	1	NEC1_RAT
33	82	6.4	774	1	AMY2_SCHPO

RESULT 1				
ID	GFP_AEQVI	STANDARD;	PRT;	238 AA.
AC	P42212; Q17104; Q27903;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent			
RT	protein."			
RL	Gene 111:229-233(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and			
RT	fluorescence characteristics of the recombinant protein.";			
RL	FEBS Lett. 341:277-280(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97299832; PubMed=9154981;			
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;			
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent			
RT	protein by modification of its codon usage.";			
RL	Plant Mol. Biol. 33:989-999(1997).			
RN	[4]			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea			
RT	green-fluorescent protein.";			
RL	Biochemistry 32:1212-1218(1993).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,			
RA	Remington S.J.;			
RT	"Crystal structure of the Aequorea victoria green fluorescent			
RT	protein.";			
RL	Science 273:1392-1395(1996).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein.";			
RL	Nat. Biotechnol. 14:1246-1251(1996).			

P49052 bacillus li
Q7vni0 haemophilus
P53621 homo sapien
Q25443 helicobacte
Q9ugn5 homo sapien
Q00116 homo sapien
Q8zdf8 yersinia pe
P91679 drosophila
P03162 duck hepati
Q9pkt7 chlamydia m
P94126 azorhizobiu
Q63416 rattus norv

ALIGNMENTS

[7] X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
 MEDLINE=98455509; PubMed=9782051;
 Wachter R.M., Eislinger M.A., Kallio K., Hanson G.T., Remington S.J.;
 "Structural basis of spectral shifts in the yellow-emission variants
 of green fluorescent protein."; Structure 6:1267-1277(1998).
 [8] X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 MEDLINE=99238303; PubMed=10220315;
 Eislinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
 "Structural and spectral response of green fluorescent protein
 variants to changes in pH."; Biochemistry 38:5296-5301(1999).
 CC -1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the
 blue chemiluminescence of the protein aequorin into green
 fluorescent light by energy transfer. Fluoresces in vivo upon
 receiving energy from the Ca(2+)-activated photoprotein aequorin.
 Absorbs light maximally at 395 nm and exhibits a smaller
 absorbance peak at 470 nm. The fluorescence emission spectrum
 peaks at 509 nm with a shoulder at 540 nm.
 CC -1- SUBUNIT: Monomer.
 CC -1- TISSUE SPECIFICITY: Photocytes.
 CC -1- PTM: Contains a covalently attached chromophore, which is composed
 of modified amino acid residues. The chromophore is formed upon
 cyclization of the residues Ser-dehydroTy-Gly.
 CC -1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
 chimeric proteins of GFP linked to other proteins where it
 functions as a fluorescent protein tag. GFP tolerates N- and C-
 terminal fusion to a broad variety of proteins. It has been
 expressed in bacteria, yeast, slime mold, plants, Drosophila,
 zebrafish, and in mammalian cells. As a noninvasive fluorescent
 marker in living cells, it allows for a wide range of applications
 where it may function as a cell lineage tracer, reporter of gene
 expression, or as a measure of protein-protein interactions.
 CC -1- DATABASE: NAMB-Protein Spotlight;
 NOT-Base Issue 11 of June 2001;
 WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; M62654; AAA27722.1; -;
 DR EMBL; M62653; AAA27721.1; -;
 DR EMBL; J29345; AAA58246.1; -;
 DR EMBL; X96418; CAA65278.1; -;
 DR PIR; J30692; JQ1514.
 DR PDB; 1B9C; 17-NOV-00.
 DR PDB; 1BFP; 07-JUL-97.
 DR PDB; 1CAF; 14-JUN-00.
 DR PDB; 1EMA; 08-NOV-96.
 DR PDB; 1EMB; 16-JUN-97.
 DR PDB; 1EMC; 20-AUG-97.
 DR PDB; 1EME; 20-AUG-97.
 DR PDB; 1EMF; 20-AUG-97.
 DR PDB; 1EMG; 12-MAY-99.
 DR PDB; 1EMK; 20-AUG-97.
 DR PDB; 1EML; 20-AUG-97.
 DR PDB; 1EMW; 20-AUG-97.
 DR PDB; 1F09; 17-NOV-00.
 DR PDB; 1FOB; 17-NOV-00.
 DR PDB; 1GFL; 11-JAN-97.
 DR PDB; 1H0J; 15-JAN-02.
 DR PDB; 1H0Y; 04-JUL-01.
 DR PDB; 1JB1; 07-JAN-03.
 DR PDB; 1JB2; 07-JAN-03.
 DR PDB; 1KP5; 28-AUG-02.
 DR PDB; 1KYP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.
 DR PDB; 1KYS; 10-APR-02.
 DR PDB; 1YFP; 28-OCT-98.
 DR PDB; 2EMD; 20-AUG-97.
 DR PDB; 2EMN; 20-AUG-97.
 DR PDB; 2EMO; 20-AUG-97.
 DR InterPro: IPR009017; GFP like.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PD01229; GFLUORESCENT.
 DR ProDom; PD013756; Green_fl_protein; 1.
 KW Luminescence; 3D-structure.
 FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).
 FT MOD_RES 66 66 2,3-DIDEHYDROTYROSINE.
 FT VARIANT 100 100 F -> Y.
 FT VARIANT 108 108 T -> S.
 FT VARIANT 141 141 L -> M.
 FT VARIANT 219 219 V -> I.
 FT CONFLICT 2 2 S -> G (IN REF. 3).
 FT CONFLICT 25 25 H -> Q (IN REF. 2).
 FT CONFLICT 80 80 Q -> R (IN REF. 3).
 FT CONFLICT 157 157 Q -> P (IN REF. 2).
 FT CONFLICT 172 172 E -> K (IN REF. 2).
 FT HELIX 4 8
 FT STRAND 12 22
 FT TURN 23 24
 FT STRAND 25 36
 FT TURN 37 40
 FT STRAND 41 48
 FT TURN 49 50
 FT TURN 57 60
 FT HELIX 61 63
 FT TURN 69 71
 FT HELIX 73 73
 FT STRAND 76 81
 FT HELIX 83 86
 FT TURN 87 90
 FT STRAND 92 100
 FT TURN 101 102
 FT STRAND 105 115
 FT TURN 116 117
 FT STRAND 118 128
 FT TURN 132 133
 FT TURN 135 139
 FT STRAND 141 141
 FT STRAND 148 155
 FT TURN 156 159
 FT STRAND 160 171
 FT TURN 172 173
 FT STRAND 176 187
 FT STRAND 199 208
 FT TURN 211 212
 FT STRAND 217 227
 SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6E05 CRC64;
 Query Match 97.6%; Score 1242; DB 1; Length 238;
 Best Local Similarity 97.9%; Pred. No. 1.7e-95;
 Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 VSKGELFTGVVPIVLVELDGDVNGHKFSVSGEGEDATYGLTLKFCITTKGLPVPWPTL 61
 Db 1 MSKGELFTGVVPIVLVELDGDVNGHKFSVSGEGEDATYGLTLKFCITTKGLPVPWPTL 60
 QY 62 VTVLVSVQCFRSYRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
 Db 61 VTTFSYGVQCFRSYRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
 QY 122 NRIELKGIIDFKEDGNILGHKLEYNVNSHNVIYIMADQKNGIKVNFKIRNIEDGSYQLAD 181
 Db 121 NRIELKGIIDFKEDGNILGHKLEYNVNSHNVIYIMADQKNGIKVNFKIRNIEDGSYQLAD 180
 QY 182 HYQONTPIGDGPEVLLPDNHYLTSQSALSADPNKRDHMYLLGLFVTAAGITLGMDELYK 239

104	QY	104	DNGYKTRAEVKEGGDTLVNRIELKGLDFKEDGNILGHKLEYNVSHVYIMADK-OKNGI	163					
365	DB	365	DEBIDLTQAFVHEHGKLVNSDFDGKNF--DGAENG-----IAKLEKIGV	408					
163	QY	163	---KYNFKIRH-----NIEDGSVQLADHYQQQNTPTICDGPVLLPNDHYL-	202					
409	DB	409	GKEQVNYRLRDWGVSRQRYWGAPIPMLTLENGDVVPA-----PWEDLPILPDEVVMD	461					
203	QY	203	STQSALSADPN 213						
462	DB	462	GVKSPINADPN 472						
RESULT 3									
SYL_XYLFA									
AC	SYL_XYLFA	STANDARD;	PRT;	879 AA.					
AC	Q9PBG8:								
DT	16-OCT-2001	(Rel. 40, Created)							
DT	16-OCT-2001	(Rel. 40, Last sequence update)							
DT	28-FEB-2003	(Rel. 41, Last annotation update)							
DT	LEUCY-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (Leurs).								
DE	LEU OR XF2176.								
GN	Xylella fastidiosa.								
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;								
OC	Xanthomonadaceae; Xylella.								
OC	NCBI_TaxID=2371;								
OX	[1]								
RN	SEQUENCE FROM N.A.								
RC	STRAIN=9a5C;								
RP	MEDLINE=20365717; PubMed=10910347;								
RX	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,								
RA	Alvarenga R., Alves L.M.C., Arya J.E., Baia G.S., Baptista C.S.,								
RA	Bartos M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,								
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,								
RA	Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,								
RA	Coutinho L.P., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,								
RA	Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,								
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,								
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,								
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,								
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,								
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,								
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,								
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,								
RA	Menck C.F.M., Mitracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,								
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,								
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,								
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,								
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,								
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,								
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,								
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,								
RA	da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,								
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,								
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,								
RA	Zago M.A., Zatz M., Meidanis J., Setubal J.C.;								
RT	"The genome sequence of the plant pathogen Xylella fastidiosa.";								
RT	Nature 406.151-159(2000)								
CC	-!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +								
CC	diphosphate + L-leucyl-tRNA(Leu).								
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.								
CC	-!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.								
CC	-----								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation								
CC	the European Bioinformatics Institute. There are no restrictions on use								
CC	by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/).								
CC	or send an email to license@isb-sib.ch).								

181 HQQNTPTGDPVLLPDHNYLSTQSALSKDPNKRDRHVVLEFVTAAGITHGMDELYK 238

DB

104 DGNKYTRAEVFEQDGTAVNRLELGDIDPKEDGNILGHKLENNVSHNYIMADK-QKNGI 163

DB

365 DEEIDLTKQAFVFEHGKLVNSDFDGKNF--DGAENG-----IADKLEKIGV 408

QY

163 ---KVNFKIRH-----NIEDGSVOLADHYQOQNTPTGDPVLLPDHNYL- 202

DB

409 GKRQVNYELRDGVSQRQYWGAPIMLTLENGDVVPA-----PMEDLPILPDEVMD 461

QY

203 STQSALSADPN 213

DB

462 GVKSPINADPN 472

RESULT 3

SYL_XYLFA

ID SYL_XYLFA STANDARD; PRT; 879 AA.

AC Q9PBG8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).

GN LEUS OR XF2176.

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xylella.

OX NCBI_TaxID=2371;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=9a5c;

RX MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,

RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,

RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,

RA Coutinho L.L., Crisofani M., Dias-Neto E., Docena C., El-Dorry H.,

RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,

RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,

RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,

RA Krieger J.E., Kurame E.E., Laigret F., Lambais M.R., Leite L.C.C.,

RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

RA Marques M.V., Martins E.M.F., Martins E.M.F., Matsukuma A.Y.,

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,

RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,

RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,

RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,

RA de Rosa V.B. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,

RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Teuhako M.H.,

RA Vialla H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Zagoda M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159(2000).

CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +

CC diphosphate + L-leucyl-tRNA(Leu).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on use

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>).

CC or send an email to license@isb-sib.ch.

CC

CC EMBL; AF004031; AAF84975.1; ALT_INIT.

DR HAMAP; MF 00049; -; 1.

DB

181 HQQNTPTGDPVLLPDHNYLSTQSALSKDPNKRDRHVVLEFVTAAGITHGMDELYK 238

DB

104 DGNKYTRAEVFEQDGTAVNRLELGDIDPKEDGNILGHKLENNVSHNYIMADK-QKNGI 163

DB

365 DEEIDLTKQAFVFEHGKLVNSDFDGKNF--DGAENG-----IADKLEKIGV 408

QY

163 ---KVNFKIRH-----NIEDGSVOLADHYQOQNTPTGDPVLLPDHNYL- 202

DB

409 GKRQVNYELRDGVSQRQYWGAPIMLTLENGDVVPA-----PMEDLPILPDEVMD 461

QY

203 STQSALSADPN 213

DB

462 GVKSPINADPN 472

RESULT 2

SYL_HAEIN

ID SYL_HAEIN STANDARD; PRT; 861 AA.

AC P43827;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).

GN LEUS OR HI0921.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=rd / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spillings T., Hedblom E., Cotton M.D.,

RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

RA "Whole-genome random sequencing and assembly of Haemophilus influenzae

RD.";

RL Science 269:496-512(1995).

CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +

CC diphosphate + L-leucyl-tRNA(Leu).

CC -!- SUBUNIT: Monomer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>).

CC or send an email to license@isb-sib.ch.

CC

CC EMBL; U32774; AAC22581.1; -.

DR PIR; H64102; H64102.

DR TIGR; HI0921; -.

DR HAMAP; MF 00049; -; 1.

DR InterPro; IPR002302; Leu-trNASyntla.

DR InterPro; IPR002300; tRNA-synt_1a.

DR InterPro; IPR001412; tRNA-synt_1.

DR InterPro; IPR009008; VALRS_ILERS_edit.

DR Pfam; PF00133; tRNA-synt 1; 1.

DR PRINTS; PR00985; TRNASYNTHLEU.

DR TIGRFamE; TIGR00396; leuS_bact; 1.

DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.

DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW Complete proteome.

FT SITE 42 52

FT SITE 619 623

FT BINDING 622 623

FT SEQUENCE 861 AA; 97750 MW; EB93304F6B4C6FB7 CRC64;

FT "HIGH" REGION.

FT "KMSKS" REGION.

FT ATP (BY SIMILARITY).

Query Match 7.3%; Score 93.5; DB 1; Length 861;

Best Local Similarity 24.1%; Pred. No. 4.5;

Matches 46; Conservative 27; Mismatches 59; Gaps 11;

50 TTGKLPVPVPTLVTLSVGQCFSRYPDHMKQHDFFKSAMPEGVQVQRTIFFKD----- 103

314 TGDKLPI-VWANFVLMHYGTGVAMVPAH-DQRDF-----EFAQKYSLPKQVIAPLA 364


```
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC di-phosphate + L-leucyl-tRNA(Leu).
CC SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE01725; AAL19599.1; --
CC STyGene; SG77777; leuS.
CC HAMAP; MF_00049; -, 1.
CC InterPro; IPR002302; Leu-tRNA-synt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS IleRS_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 42 52 "HIGH" REGION.
CC SITE 619 623 "KMSKS" REGION.
CC BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;
CC -----
Query Match 7.0%; Score 89.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 9.6;
Matches 42; Conservative 22; Mismatches 77; Indels 39; Gaps 7;
QY 50 TTGKLPVPPTLVTLVLSYGVCFSRYPDHMKQHDFFKSAFPGVGH-QDRD-YEFASKYGLTIKPVLAADGSEPD 109
DB 314 TGEIIPV-WAANFVLMYGTGAVMVGPH-QDRD-YEFASKYGLTIKPVLAADGSEPD 370
QY 110 RAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNHNVIMADKQKNGIKVNFKR 169
DB 371 SEQALTEKGVLFNSGFDGLAFEAFAFNAIDKL-----AEKGVGERGVNRLR 418
QY 170 H-----NIEDGSVOLADHYQNTPIGDGVLDPDNHNL-STQSALSNDP 212
DB 419 DWGVSQRVWAPIMVTLEDGTV-----LPTPEDQLPVLPEDVMDGITSPKADP 471
RESULT 6
SYL_XYLF
ID SYL_XYLF STANDARD; PRT; 879 AA.
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR PD1230.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carlier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan P.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
```

```
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Takumoto F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa."
RL J. Bacteriol. 185:1018-1026(2003).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC di-phosphate + L-leucyl-tRNA(Leu).
CC SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE012557; AAO29080.1; ALT_INIT.
CC HAMAP; MF_00049; -, 1.
CC InterPro; IPR002302; Leu-tRNA-synt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS IleRS_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 45 55 "HIGH" REGION.
CC SITE 637 641 "KMSKS" REGION.
CC BINDING 640 640 ATP (BY SIMILARITY).
CC SEQUENCE 879 AA; 99823 MW; 4C2EE01B8FDC497E CRC64;
CC -----
Query Match 7.0%; Score 88.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 12;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;
QY 50 TTGKLPVPPTLVTLVLSYGVCFSRYPDHMKQHDFFKSAFPGVGH-QDRD-YEFASKYGLTIKPVLAADGSEPD 107
DB 321 TNEQLPV-WVANFVLMYGTGAVMVGPH-QDRD-YEFASKYGLTIKPVLAADGSEPD 377
QY 108 -----KTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNHNVIMADKQKNGIKVNFKR 153
DB 378 STWEPDVRDWDYADKTR---EFE---LINSABFDGLDYQCAFEVLAERFE----- 421
QY 154 MADKQKNG-ISKVNFKIRHNIEDGSVOLADHYQNTPI-----GDGVPVLLPDN 199
DB 422 ---RQGRQRRVNYRLR---DWGVSQRVWAPIMVTLEDGTV-----LPTPEDQLPVLPEDVMDGITSPKADP 471
QY 200 -HYLSTQSALSNDPNEKR 216
DB 475 VAFSGTGTSPKIDPEWRK 492
RESULT 7
CP51_CANGA
ID CP51_CANGA STANDARD; PRT; 533 AA.
AC F50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CyP51) (P450-LIAl) (Sterol 14-
DE alpha-demethylase) (lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
```

```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=2001-L5;
RX  MEDLINE=96161286; PubMed=8593007;
RA  Geber A., Hitchcock C.A., Swartz J.B., Pullen F.S., Marsden K.E.,
RA  Kwon-Chung K.J., Bennett J.E.;
RT  "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT  viability, cell growth, sterol composition, and antifungal
RT  susceptibility.";
RL  Antimicrob. Agents Chemother. 39:2708-2717(1995).
[2]
RN  SEQUENCE OF 60-473 FROM N.A.
RP  STRAIN=ATCC 2001;
RX  MEDLINE=95081364; PubMed=7989540;
RA  Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA  Rossier M.;
RT  "Rapid detection and identification of Candida albicans and
RT  Torulopsis (Candida) glabrata in clinical specimens by
RT  species-specific nested PCR amplification of a cytochrome P-450
RT  lanosterol-alpha-demethylase (LAL1) gene fragment.";
RL  J. Clin. Microbiol. 32:1902-1907(1994).
CC  -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC  critical for ergosterol biosynthesis. It transforms lanosterol
CC  into 4,4'-dimethyl cholesterol-8,14,24-triene-3-beta-ol (By
CC  similarity).
CC  -!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC  methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC  NADP(+) + 3 H(2)O.
CC  -!- PATHWAY: Ergosterol biosynthesis.
CC  -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; U40389; AAB02329.1; -.
DR  EMBL; S75389; AAB32679.1; -.
DR  InterPro; IPR001128; Cytochrome_P450.
DR  Pfam; PF00067; P450.1.
DR  PRINTS; PR00385; P450.
DR  PROSITE; PS00086; CYTOCHROME_P450; 1.
KW  Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
KW  Sterol biosynthesis; NADP.
FT  METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT  CONFLICT 64 64 I -> M (IN REF. 2).
FT  CONFLICT 473 473 I -> T (IN REF. 2).
SQ  SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.8%; Pred. No. 9.6;
Matches 45; Conservative 32; Mismatches 80; Indels 49; Gaps 9;

QY 25 GHKFSVS---GEGEGATYCKILKICITCKLPVWPVLTVTLVLSGVQCFSRYPDH--M 79
Db 109 GHEFIFNAKLADVSAEAAYSHLT-----TPV-----FGKGIYDCPNHRLM 149
QY 30 KQHDFFKSAM-PRGYV-----QERTIFFKDDGNKYKRAEVKFGDGLVNRILKGDIF 131
Db 150 EQKKFKVKGALTKFAFVRYVPLIAEEYKYFRNSKNFKINENSGIVDVVVSQPEM--TTF 207
QY 132 KEDGNILGHKLEYNYNHNVNIMADKQKGIKYNFKIRNIEDGVSQVLADHYQNTPIGD 191
Db 208 TASRSLGLKEMRKLDTFDAFLYSLDLDKGTPTINF-VFENLPLEHYVKRDKDHAQAIS--- 263
QY 132 GPVLLPDNHYLSTQSALSCKDPEKRD 217
Db 254 -----GTYSMLIKERRKND 278
```

```

RESULT 8
DPOL_HPBHE STANDARD; PRT; 788 AA.
AC P13846;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P protein [Includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-
DE directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].
GN P.
OS Heron hepatitis b virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=28300;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88333160; PubMed=3418788;
RA Spengel R., Kaleta E.F., Will H.;
RT "Isolation and characterization of a hepatitis B virus endemic in
RT herons.";
RL J. Virol. 62:3832-3839(1988).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}[N].
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR  EMBL; M22056; AAA45738.1; -.
DR  PIR; A30082; JDVLHH.
DR  InterPro; IPR001462; DNAPol_viral_C.
DR  InterPro; IPR000201; DNAPol_viral_N.
DR  InterPro; IPR000477; RVTse.
DR  Pfam; PF00336; DNA_pol_viral_C; 1.
DR  Pfam; PF00242; DNA_pol_viral_N; 1.
DR  Pfam; PF00078; rvt; 1.
DR  ProDom; PD000814; DNAPol_viral_C; 1.
DR  ProDom; PD000814; DNAPol_viral_C; 1.
KW  Transferase; RNA-directed DNA polymerase; DNA-directed DNA polymerase;
KW  Hydrolase; Nuclease; Endonuclease; DNA replication; DNA-binding.
SQ  SEQUENCE 788 AA; 90070 MW; FB44F38F75EADF44 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 788;
Best Local Similarity 19.7%; Pred. No. 15;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;

QY 58 WPTLVTVLSYGVQCFSRYPDHMKQH-----DFKSAPEGYVQERT----IPFKDDGNKY 108
Db 139 WPKSISYLPVHSGVKPKYPEFOONHESLVNDYLNKLFEAGILYKRVSKHLVTFK--GPVF 196
QY 109 T-----RAEVKFGDGLVNRILKGDIFKEDGNILGHKLEYNYNHNVNIM 154
Db 197 TWEQKHLVPOQHGAYSSKINDQESRRRIITATSRKNDSSRI-----FGAHN---- 245
QY 155 ADKQKNGIKVNFKIRNIEDGVSQVLADHYQ-----QNTPIGDGPVLL--PNHYL 202
Db 246 -----NGRKISY---HSTRDGSRLSGRTSDPTRSGALAGGSDTPIGPGSTAHPSTHHV 297
QY 203 STQ-----SALSKDPNEKR 216
Db 298 DRRRRQKGGVQLQAISREPSETR 320

RESULT 9
D152_HAEIN
ID D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
```

RESULT 10
D151_HAEIN
ID_D151_
AC P4602
DT 01-N
DT 01-N
DT 16-O
DE Prote

DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR S01174.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OX Alteromonadaceae; Shewanella.
RN NCBI_TaxID=70863;
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
R.A. Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,
Vamathavan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB015561; AN54244.1; -.
DR TIGR; S01174; -.
DR HAMAP; MF 00049; -; 1.
DR InterPro; IPR002302; Leu-tRNA-synt1a.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_fIleRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 618 622 "KMSKS" REGION.
FT BINDING 621 621 ATP (BY SIMILARITY).
SQ SEQUENCE 859 AA; 96827 MW; C78D6209DFB6CA17 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 859;
Best Local Similarity 23.6%; Pred. No. 17;
Matches 50; Conservative 30; Mismatches 95; Indels 37; Gaps 9;

QY 18 ELDDGVNGHKFSVSGE-----GDATVGKLTLCFICTGKLPVWPPLTVLVSQVQ 70
DB 277 ELAATFDECKNSTSRAELATMEKRGVAT-GLYAIHPI--TGQVPIWAANFVLMNYGTG 333
QY 71 CFSRYPDH-MQHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDPLVNRIBLKG 129
DB 334 AVMSVPGHQDQDVEFAK----KVHLPTFAVIKPAEGDLDTSEAAYTEKGLIFNSGEFDGL 389
QY 130 DPKEDGNILGHKLEYNSHNHYIMADKQNGIKVNFKIRH-----NIEDGSGVQLA 180
DB 390 DFDGAFNFIANKL-----VAEGKGR--QVNYRLRDMGVSRQRYWGAPIPMVTLA 437
QY 131 DHYQONTPIGDGVPVLLPDNHYL-STOSALSKD 211
DB 438 DGTVIPTPEDQLPVLPEDVMDGIQSPIKAD 469

RESULT 12
TRNA_CAMJE STANDARD; PRT; 357 AA.
AC Q9PP92; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5-)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-
methyltransferase) (RUMT).
DE TRMA OR C30831C.
GN Campylobacter jejuni.
OS Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteriaceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
54 (M-5-U54) in all tRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
homocysteine + tRNA containing thymine.
CC -1- SIMILARITY: Belongs to the RNA MSU methyltransferase family. Trna
subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL139076; CAB73096.1; -.
DR PIR; G81355; G81355.
DR HAMAP; MF 01011; -; 1.
DR InterPro; IPR000051; SAM bind.
DR InterPro; IPR001566; TrmA.
DR PROSITE; PS01230; TRMA_1; 1.
DR PROSITE; PS01231; TRMA_2; FALSE NEG.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY
SIMILARITY).
FT ACT_SITE 315 315 BY SIMILARITY.
SQ SEQUENCE 357 AA; 42276 MW; CEC5328347CEB497 CRC64;

Query Match 6.8%; Score 86; DB 1; Length 357;
Best Local Similarity 24.8%; Pred. No. 6.5;
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;

QY 80 KQHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKF--EGDPLV-----NRLEK 128
DB 14 EKHSITKIKYKFFYTKDFKLFASKDKHVKTRAEVLSFYHENDTLFYAMFDPKSKKYLIE 73
QY 129 IDPKED-----GNILGHKLEYNSHNHYIMADKQNGIKVNFKIRHIE 173
DB 74 LQFADEKICAFPRLLLEYLRQDNKLEKL-----FGVEFLTTKQE--LSITLLYHK 125
QY 174 D 174
DB 126 D 126

RESULT 13
PSBC_CYAPA

	Matches	40; Conservative	34; Mismatches	70; Indels	67; Gaps	9;
Qy	41	GKLTLLKFICT-TGKLPVPMPTLVT	VLVSYGVCFSY--PDHMKQHDFFKSAMPEGYQVR	97		
Db	313	GRDLKMLCAEDQSRMCWTAIRLLKYGMQYQNYMHP	SQARSACSQSPMSVSEN	372		
Qy	98	TIFFKDDGNYKTRAEVKFEGDTLVNRIELKIDF	KEDG-----NIL	138		
Db	373	SLVAMDFSGKTRV-IDNPTEALSAVE-EGLAWRKGLCLRLGNHGS	PAPQSSAVNNA	430		
Qy	139	GKLEYNYNHSH-----NVYIMADKQKN-----	GKVNFE-----166			
Db	431	LHRSQWPFHHRISRDEAQQLITRQGVGVFLVRDSQSNPRTFVLSMHSKGQIKHFIIP	490			
Qy	167	-----KIRHNIEDGS-----VOLADHYQON	186			
Db	491	VEDDGEVFTLDDGHTKFTDILQLVFEYQLN	521			

RESULT 15
D153 HAEIN
ID D153 HAEIN STANDARD; PRT; 793 AA.
AC Q32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RX MEDLINE=97427952; PubMed=9284140;
RA Loomscre S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: U60834; AAB61977.1; -;
InterPro: IPR000184; Bac_surfAg_D15.
Pfam: PF01103; Bac_surface_Ag; 1.
Kw Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A14 CRC64;

Query Match 6.7%; Score 85.5; DB 1; Length 793;
Best Local Similarity 22.4%; Pred. No. 19;
Matches 49; Conservative 28; Mismatches 79; Indels 63; Gaps 11;

QY	65	LSYGVCQFSRY	PDHMKQHDFF	FKSAMPEGYQVR	-----	RTI	99
Db	427	IGYTESGISY	QTSIKQDNFLGT	GAVSIAGTKNDY	GSVNLGYTE	FTPKDGVSLGNI	486
QY	100	FFKDDGNYK	TRAEVKFEGDTL	VNRIELKIDF	KEDGNI---	LGH-----	KLEYNYS 148
Db	487	FFENVDSKSD	TSSNYKRTTYG	SNVTL-GFP	VNNNSYYVGL	GHTYKNKISN	FALEYN--- 542

QY	149	HNVTIMADKQK	-NGIKVNFKIR	HNIEDGSVQL	ADHYQQ-----	NTPIGDPVLL	196
Db	543	RNLYIQSMK	FKGNGIKTN	-----	DFDFSFGWNY	NSLNRGYPTK	GVKASLG-GRVTI 593
QY	197	P--DNHYLS	TQSALSKDP	NEKRDHML	LVLLGFVTA	AGITLG	233
Db	594	FGSDNKYYK	LSADVQGF	YPLDRDR	HRWVVS	AKASAGYANG	632

Search completed: June 21, 2004, 15:55:20
Job time : 7.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds
(without alignments)
2458.984 Million cell updates/sec

Title: US-09-887-784-64v
Perfect score: 1273
Sequence: 1 MNSGSELFTGVVPIVLVD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_25:*
- 1: sp archaea:*
 - 2: sp bacteria:*
 - 3: sp fungi:*
 - 4: sp human:*
 - 5: sp invertebrate:*
 - 6: sp mammal:*
 - 7: sp mhc:*
 - 8: sp organelle:*
 - 9: sp phage:*
 - 10: sp plant:*
 - 11: sp rodent:*
 - 12: sp virus:*
 - 13: sp vertebrate:*
 - 14: sp unclassified:*
 - 15: sp virus:*
 - 16: sp bacteriaph:*
 - 17: sp archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1238	97.3	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1235	97.0	238	5 Q93125	Q93125 aequorea vi
3	1233	96.9	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1232	96.8	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1200	94.3	238	5 Q17105	Q17105 aequorea vi
6	1185	93.1	238	5 Q17106	Q17106 aequorea vi
7	1080	84.8	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1076	84.5	238	5 Q8WTP95	Q8wtp95 aequorea ma
9	1072	84.2	238	5 Q8WTC4	Q8wtc4 aequorea ma
10	1070	84.1	238	5 Q8WTD0	Q8wtc0 aequorea ma
11	1069	84.0	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1069	84.0	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1067	83.8	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1065	83.7	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5 Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5 Q7Z0W5	Q7z0w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	246.5	19.4	225	5	Q7Z0W9	Q7z0w9 montastraea
19	244.5	19.2	236	5	Q8T6U0	Q8t6u0 dendronephth
20	240	18.9	225	5	Q8I6J8	Q8i6j8 trachyphyll
21	238.5	18.7	286	5	Q9U6Y3	Q9u6y3 clavularia
22	236	18.5	225	5	Q8T5F1	Q8t5f1 montastraea
23	233	18.3	225	5	Q7Z0W4	Q7z0w4 montastraea
24	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
25	218	17.1	227	5	Q7Z0W6	Q7z0w6 montastraea
26	218	17.1	234	5	Q7Z0W7	Q7z0w7 montastraea
27	213	16.7	234	5	Q8T5F2	Q8t5f2 montastraea
28	213	16.7	234	5	Q8MU47	Q8mu47 montastraea
29	212.5	16.7	229	5	Q9U6Y6	Q9u6y6 anemonia ma
30	212.5	16.7	259	5	Q8MMA2	Q8mma2 agaricia fr
31	212	16.7	239	5	Q8MMA1	Q8mma1 agaricia ag
32	210	16.5	227	5	Q962P9	Q962p9 montastraea
33	210	16.5	227	5	Q7Z0W8	Q7z0w8 montastraea
34	206	16.2	227	5	Q95VT0	Q95vt0 montastraea
35	203.5	16.0	238	5	Q9BLY9	Q9bly9 renilla mue
36	203	15.9	221	5	Q95P04	Q95p04 goniopora t
37	202.5	15.9	232	5	Q9GP15	Q9gp15 anemonia s
38	200.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
39	199.5	15.7	214	5	Q86LV7	Q86lv7 meandrina m
40	198.5	15.6	214	5	Q86LV8	Q86lv8 meandrina m
41	198.5	15.6	222	5	Q7Z168	Q7z168 cerianthus
42	198.5	15.6	225	5	Q8T6T9	Q8t6t9 radianthus
43	198.5	15.6	232	5	Q9GZ28	Q9gz28 anemonia su
44	198	15.6	235	5	Q8T5F0	Q8t5f0 scolymia cu
45	196.5	15.4	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	ID	Q8GHE2	PRELIMINARY;	PRT;	238 AA.
AC	Q8GHE2;				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Green fluorescence protein.				
GN	2289GFP.				
OS	Azotobacter vinelandii.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Azotobacter.				
OX	NCBI_TaxID=354;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DSM2289;				
RA	Koranyi P., Berenyi M., Burg K.;				
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria				
RT	Azomonas and Azotobacter.";				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF324408; AAN86140.1; -;				
DR	GO; GO:0006091; P:energy pathways; IEA.				
DR	InterPro; IPR009017; GFP_like.				
DR	InterPro; IPR000786; Green_fl_protein.				
DR	Pfam; PF01353; GFP; 1.				
DR	PRINTS; PR01229; GFP_LUORESCENT.				
DR	ProDom; PD013756; Green fl protein; 1.				
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;				

Query Match 97.3%; Score 1238; DB 2; Length 238;
Best Local Similarity 97.5%; Pred. No. 9.3e-97;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy	2	VSKGEELFTGVVPIVLVDGNGHKFSVSGEGEDATYGLTKLTKFTCTTGKLPVWPPTL	61
Db	1	MSGGEELFTGVVPIVLVDGNGHKFSVSGEGEDATYGLTKLTKFTCTTGKLPVWPPTL	60
Qy	62	VTVLSYGVOCFSRYPDHMKOHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEFGDTLV	121

```
Db 61 VTTFSYGVQCFSRYPDHNKRDHDFKSAPEGYVQERTIFFKDDGNKYTRAEVKEGDTLV 120
Qy 122 NRIELKGIIDPKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIIDPKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 180
Qy 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 238

RESULT 2
Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RA "Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSP; P42212; IBBP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green fl protein; 1.
DR SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Qy 2 VSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPTL 60

Qy 62 VTVLSYGVQCFSRYPDHNKRDHDFKSAPEGYVQERTIFFKDDGNKYTRAEVKEGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHNKRDHDFKSAPEGYVQERTIFFKDDGNKYTRAEVKEGDTLV 120

Qy 122 NRIELKGIIDPKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIIDPKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 180

Qy 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 238

RESULT 3
Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Green fluorescent protein.
GN 8SGFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green fl protein; 1.
DR SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Qy 2 VSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPTL 60

Qy 62 VTVLSYGVQCFSRYPDHNKRDHDFKSAPEGYVQERTIFFKDDGNKYTRAEVKEGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHNKRDHDFKSAPEGYVQERTIFFKDDGNKYTRAEVKEGDTLV 120

Qy 122 NRIELKGIIDPKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIIDPKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 180

Qy 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 238

RESULT 4
Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN 8SGFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green fl protein; 1.
DR SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.8%; Score 1232; DB 2; Length 238;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
Best Local Similarity 97.1%; Pred. No. 3e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKLTILKFTCTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKLTILKFTCTTGKLPVWPPTL 60

QY 62 VTVLISGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKYVFKIRHNIEDGSVQLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKYVFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58790.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; IGFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748544 CRC64;

Query Match 94.3%; Score 1200; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 1.5e-93;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKLTILKFTCTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKLTILKFTCTTGKLPVWPPTL 60

QY 62 VTVLISGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKYVFKIRHNIEDGSVQLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKYVFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
```

```
AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; IBFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;

Query Match 93.1%; Score 1185; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 2.8e-92;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKLTILKFTCTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKLTILKFTCTTGKLPVWPPTL 60

QY 62 VTVLISGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKYVFKIRHNIEDGSVQLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKYVFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL "Colorful mutants of green fluorescent protein from Aequorea
RL macrodactyla.";
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
```

```
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 688FD75E88926903 CRC64;

Query Match      84.8%; Score 1080; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 2.2e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGIVPVLIELDGVVHGHKFSVRGEGDADYKGLKFKICTTGKLPVWPPTL 60

QY 52 VTVLSGVQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 51 VTTFSGVIGCFARYPEHMKWDDFFKSAMPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIDFKEDGNILGHKLEYNHSHVNYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRLEKGMDFKEDGNILGHKLEYNHSHVNYIMPDKANGLKVNFKIRHNIEGGVOLAD 180

QY 132 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 131 HYQTNVPLDGPVLIPINHVLSQTQTAISKDRNETRDHVMVLEFFSACGHTHGMDELYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXW.

OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPx161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.2%; Score 1072; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGIVPVLIELDGVVHGHKFSVRGEGDADYKGLKFKICTTGKLPVWPPTL 60

QY 62 VTVLSGVQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSGVIGCFARYPEHMKWDDFFKSAMPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIDFKEDGNILGHKLEYNHSHVNYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRLEKGMDFKEDGNILGHKLEYNHSHVNYIMPDKANGLKVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLDGPVLIPINHVLSQTQTAISKDRNETRDHVMVLEFFSACGHTHGMDELYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.

OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPx161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.5%; Score 1076; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 4.7e-83;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGIVPVLIELDGVVHGHKFSVRGEGDADYKGLKFKICTTGKLPVWPPTL 60

QY 52 VTVLSGVQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 51 VTTFSGVIGCFARYPEHMKWDDFFKSAMPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIDFKEDGNILGHKLEYNHSHVNYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRLEKGMDFKEDGNILGHKLEYNHSHVNYIMPDKANGLKVNFKIRHNIEGGVOLAD 180

QY 132 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 131 HYQTNVPLDGPVLIPINHVLSQTQTAISKDRNETRDHVMVLEFFSACGHTHGMDELYK 238
```

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435427; AAL33912.1; -;
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PD01229; GFP.1.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 84.0%; Score 1070; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.8e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 61

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

1 MSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 60

QY 62 VTVLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTKRAEVKFEGLTLV 121

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

61 VTTLVGVIQCFARYPEHMKMNDFFKSAMPEGYIQTERTIFFQDDGKYKTRGEVKEGDTLV 120

QY 122 NRTELKGDIPKEDGNILGHKLEYNHSHVYIMADKQKGIKYNFKIRHNIEGSGVOLAD 181

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

121 NRTELKGMDFKEDGNILGHKLEYNHSHVYIMADKQKGIKYNFKIRHNIEGSGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELK 239

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

181 HYQTNVPLGDGPVLLPINHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELK 238

RESULT 11

Q8WTC8 PRELIMINARY; PRT; 238 AA.

AC Q8WTC8; (T-EMBLrel. 20, Created)

DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea macrodactyla.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=147615;

RN [1]

RP SEQUENCE FROM N.A.

RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,

RA Li S.J., Xia N.S.;

RT "Colorful mutants of green fluorescent protein from Aequorea

RT macrodactyla.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435429; AAL33914.1; -;

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PD01229; GFP.1.

DR ProDom; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 84.0%; Score 1069; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.8e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 61

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

1 MSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 60

QY 62 VTVLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTKRAEVKFEGLTLV 121

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

61 VTTLVGVIQCFARYPEHMKMNDFFKSAMPEGYIQTERTIFFQDDGKYKTRGEVKEGDTLV 120

RESULT 12

Q8WTC9 PRELIMINARY; PRT; 238 AA.

AC Q8WTC9; (T-EMBLrel. 20, Created)

DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea macrodactyla.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=147615;

QY 122 NRTELKGDIPKEDGNILGHKLEYNHSHVYIMADKQKGIKYNFKIRHNIEGSGVOLAD 181

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

121 NRTELKGMDFKEDGNILGHKLEYNHSHVYIMADKQKGIKYNFKIRHNIEGSGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELK 239

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

181 HYQTNVPLGDGPVLLPINHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELK 238

RESULT 13

Q8WTC7 PRELIMINARY; PRT; 238 AA.

AC Q8WTC7; (T-EMBLrel. 20, Created)

DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea macrodactyla.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=147615;

RN [1]

RP SEQUENCE FROM N.A.

RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,

RA Li S.J., Xia N.S.;

RT "Colorful mutants of green fluorescent protein from Aequorea

RT macrodactyla.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435428; AAL33913.1; -;

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PD01229; GFP.1.

DR ProDom; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match 84.0%; Score 1069; DB 5; Length 238;

Best Local Similarity 81.1%; Pred. No. 1.8e-82;

Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 61

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

1 MSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 60

QY 62 VTVLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTKRAEVKFEGLTLV 121

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

61 VTTLVGVIQCFARYPEHMKMNDFFKSAMPEGYIQTERTIFFQDDGKYKTRGEVKEGDTLV 120

QY 122 NRTELKGDIPKEDGNILGHKLEYNHSHVYIMADKQKGIKYNFKIRHNIEGSGVOLAD 181

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

121 NRTELKGMDFKEDGNILGHKLEYNHSHVYIMADKQKGIKYNFKIRHNIEGSGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELK 239

Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

181 HYQTNVPLGDGPVLLPINHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELK 238

RESULT 13

Q8WTC7 PRELIMINARY; PRT; 238 AA.

AC Q8WTC7; (T-EMBLrel. 20, Created)

DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea macrodactyla.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=147615;

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds
(without alignments)
1433.395 Million cell updates/sec

Title: US-09-887-784-222A

Perfect score: 1274

Sequence: 1 MVSKGEELFTGVVPLVELD.....VLAGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1269	99.6	239	5 AAE17518	Enhanced
2	1269	99.6	363	6 ABR40352	Human ami
3	1269	99.6	893	4 AAG65781	Amino aci
4	1269	99.6	1132	4 AAG65782	Amino aci
5	1261	99.0	239	5 AAE17517	Enhanced
6	1258	98.7	239	3 AAB22882	Enhanced
7	1258	98.7	239	3 AAY54349	Amino aci
8	1258	98.7	239	3 AAY79584	EGFP sign
9	1258	98.7	239	4 AAB50804	Jellyfish
10	1258	98.7	239	4 AAB85900	A. victor
11	1258	98.7	239	4 AAB31171	Amino aci
12	1258	98.7	239	5 AAG66198	A. victor
13	1258	98.7	239	5 AAG94444	Protease
14	1258	98.7	239	5 AAE14599	Aequorea
15	1258	98.7	239	6 AAG34958	Aequorea
16	1258	98.7	239	6 AAG79829	Green flu
17	1258	98.7	239	6 ABR83616	Green flu
18	1258	98.7	239	6 ADA38074	Aequorea
19	1258	98.7	239	7 ABU63204	Aequorea
20	1258	98.7	239	7 ADC18358	EGFP (enh
21	1258	98.7	239	7 ABW00914	Aequorea
22	1258	98.7	239	7 ADE28570	Enhanced
23	1258	98.7	246	7 ABM79011	Enhanced
24	1258	98.7	248	5 AAG68319	Jellyfish
25	1258	98.7	259	5 AAU99804	Biomembra

ALIGNMENTS

RESULT 1

AAE17518

ID AAE17518 standard; protein; 239 AA.

XX AC AAE17518;

XX AC

DT 22-APR-2002 (first entry)

XX

DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.

XX

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;

KW

cellular function; genetic reporter; mutant; Stoke's shift; muten.

XX

OS Aequorea victoria.

OS

Synthetic.

XX

FH Key Location/Qualifiers

FT

Misc-difference 65

/note= "Wild type Phe substituted with Leu; This corresponds to position 64 in the wild type protein"

FT

Misc-difference 223

/note= "Wild type Glu substituted with Gly; This corresponds to position 222 in the wild type protein"

FT

WO200198338-A2.

XX

PN

27-DEC-2001.

PD

18-JUN-2001; 2001WO-EF006848.

XX

PR 19-JUN-2000; 2000DK-00000953.

PR

20-JUN-2000; 2000US-0212681P.

PR

10-MAY-2001; 2001DK-00000739.

PR

10-MAY-2001; 2001US-0290170P.

XX

(BIOI-) BIOIMAGE AS.

XX

PI Bjorn SP, Pagliaro L, Thastrup O;

XX

WPI; 2002-098224/13.

DR

N-PSDB; AAD28163.

XX

Novel fluorescent protein in in vitro assay for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution, has a green fluorescent protein with F64L and E222G mutation.

XX

Claim 9; Page 37; 41pp; English.

Aaw97451 Wild-type
Aau99803 Biomembra
Aau99802 Biomembra
Aau99800 Biomembra
Aau99801 Biomembra
Aau99807 Biomembra
Aay50142 Green flu
Aab24252 EGFP-MODC
Aau10888 EGFP-MODC
Ade28562 EGFP/ hum
Ade28564 EGFP/ hum
Ade28568 EGFP/ hum
Ade28566 EGFP/ hum
Aab22860 GFP-DEVD-
Aay79638 Caspase-3
Abg94422 Recombina
Aay42181 EGFP/DRM
Abr3620 HUB1-GFP
Aay54359 GFP mutan
Abr83621 RUB1-GFP

26 1258 98.7 265 2 AAW97451
27 1258 98.7 268 5 AAU99803
28 1258 98.7 270 5 AAU99802
29 1258 98.7 272 5 AAU99800
30 1258 98.7 273 5 AAU99801
31 1258 98.7 280 5 AAU99807
32 1258 98.7 281 3 AAY50142
33 1258 98.7 281 3 AAB24252
34 1258 98.7 281 5 AAU10888
35 1258 98.7 286 7 ADE28562
36 1258 98.7 289 7 ADE28564
37 1258 98.7 290 7 ADE28568
38 1258 98.7 290 7 ADE28566
39 1258 98.7 294 3 AAB22860
40 1258 98.7 294 3 AAY79638
41 1258 98.7 294 5 ABG94422
42 1258 98.7 308 2 AAY42181
43 1258 98.7 320 6 ABR3620
44 1258 98.7 323 3 AAY54359
45 1258 98.7 323 6 ABR83621

```

XX CC The invention relates to a fluorescent protein derived from green
XX CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
XX CC at F64L and E222G has a bigger compared to other GFP's making it very
XX CC suitable for high throughput screening due to better resolution. The
XX CC fluorescent protein is useful in vitro assays for measuring protein
XX CC kinase activity or dephosphorylation activity, or for measuring protein
XX CC redistribution. The fluorescent protein is useful in studying cellular
XX CC functions in living cells; as protein tags in transgenic animals, living
XX CC and fixed cells; organelle tags, secretion marker and genetic reporter.
XX CC The fluorescent protein is also useful as a cell or organelle integrity
XX CC marker, a marker for changes in cell morphology, as transfection marker,
XX CC and as a marker to be used in combination with fluorescence activated
XX CC cell sorting (FACS). The novel proteins can also be used as reporters to
XX CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
XX CC protein is also useful as markers in transcriptional and translational
XX CC fusions for performing transposon vector mutagenesis and as a reporter
XX CC for bacterial detection. Transposons encoding the fluorescent protein are
XX CC useful for screening promoters and for tagging plasmids and chromosomes.
XX CC The fluorescent protein engineered into the genome of a phage is useful
XX CC for designing diagnostic tool. The present sequence is a DNA encoding
XX CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
XX CC
XX CC Sequence 239 AA;
SQ
Query Match 99.6%; Score 1269; DB 5; Length 239;
Best Local Similarity 99.8%; Pred. No. 4.6e-122;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKGSEELFTGVVPIILVELDGDVNGHKFVSVSGBEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKGSEELFTGVVPIILVELDGDVNGHKFVSVSGBEGDATYVKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 2
ABR40352
ID ABR40352 standard; protein; 363 AA.
XX
AC ABR40352;
XX
DT 08-JUL-2003 (first entry)
XX
DE Human amino acid sequence SEQ ID NO: 6.
XX
DE Human; heterologous conjugate; intracellular protein.
XX
OS Homo sapiens.
XX OS Aequoria victoria.
XX
PN WO2003029827-A2.
XX
PD 10-APR-2003.
XX
PF 01-OCT-2002; 2002WO-DK000651.
XX
PR 01-OCT-2001; 2001DK-00001433.
XX
PR 11-OCT-2001; 2001US-032896P.
XX
XX (BIOL-) BIOLMAGE AS.
XX
XX Terry BR, Nielsen SJ;
XX
XX

```

```

XX DR WPI: 2003-430211/40.
XX DR N-PSDB; ACC72604.
XX
PT Novel cell for identifying modulators of protein interaction, contains a
PT first conjugate comprising anchor protein, second conjugate having type B
PT interactor protein and third conjugate with detectable group.
XX
PS Disclosure; Page 112-113; 118pp; English.
XX
XX The invention relates to a novel cell, comprising three heterologous
XX conjugates (HC), a first HC (HC1) comprising an anchor protein that
XX specifically binds to an internal structure within the cell conjugated to
XX an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
XX type B conjugated to a first protein of interest, and a third HC (HC3)
XX comprising a second protein of interest conjugated to detectable group.
XX The cell is useful for detecting if a compound disrupts or induces the
XX interaction between two intracellular proteins. The cell is also useful
XX for screening compounds that modulate the interaction between two
XX intracellular proteins. The present sequence is used in the
XX exemplification of the invention
XX
XX Sequence 363 AA;
SQ
Query Match 99.6%; Score 1269; DB 6; Length 363;
Best Local Similarity 99.6%; Pred. No. 8.6e-122;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKGSEELFTGVVPIILVELDGDVNGHKFVSVSGBEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKGSEELFTGVVPIILVELDGDVNGHKFVSVSGBEGDATYVKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 3
AAG65781
ID AAG65781 standard; protein; 893 AA.
XX
AC AAG65781;
XX
DT 07-JAN-2002 (first entry)
XX
DE Amino acid sequence of HSPDE4A1-E222G fusion protein.
XX
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX fusion protein.
XX
OS Homo sapiens.
XX OS Aequorea victoria.
XX
PN WO200179526-A2.
XX
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-DK000264.
XX
PR 17-APR-2000; 2000DK-00000651.
XX
PR 23-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX

```



```
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
DR WPI; 2001-611727/70.
DR N-PSDB; AAI66852.
XX
PT Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
PS Example 1; Page 156-160; 160pp; English.
XX
CC The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
CC protein
CC
XX Sequence 893 AA;
XX
Query Match 99.6%; Score 1269; DB 4; Length 893;
Best Local Similarity 99.6%; Pred. No. 3.2e-121;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MYSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTFLKFICTTGKLPVPWPT 60
DB 655 MYSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTFLKFICTTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTTAAAGITLGMDELYK 239
DB 835 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTTAAAGITLGMDELYK 893
RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA.
XX
AC AAG65782;
XX
DT 07-JAN-2002 (first entry)
XX
DE Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW fusion protein.
XX
OS Homo sapiens.
OS Aequorea victoria.
XX
PN WO200179526-A2.
XX
```

```
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-DK000264.
XX
PR 17-APR-2000; 2000DK-00000651.
PR 29-MAY-2000; 2000DK-00000849.
XX
PA (BIOI-) BIOIMAGE AS.
XX
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
DR WPI; 2001-611727/70.
DR N-PSDB; AAI66853.
XX
PT Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
PS Example 1; Page 162-167; 160pp; English.
XX
CC The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
CC protein
CC
XX Sequence 1132 AA;
XX
Query Match 99.6%; Score 1269; DB 4; Length 1132;
Best Local Similarity 99.6%; Pred. No. 4.5e-121;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MYSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTFLKFICTTGKLPVPWPT 60
DB 894 MYSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTFLKFICTTGKLPVPWPT 953
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTTAAAGITLGMDELYK 239
DB 1074 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTTAAAGITLGMDELYK 1132
RESULT 5
AAE17517
ID AAE17517 standard; protein; 239 AA.
XX
AC AAE17517;
XX
DT 22-APR-2002 (first entry)
XX
DE Enhanced F64L jellyfish green fluorescent protein mutant.
XX
```

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
 KW cellular function; genetic reporter; mutant; Stoke's shift; mitein.
 XX
 OS Aequorea victoria.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 65 /notes "Wild type Phe substituted with Leu; This
 FT corresponds to position 64 in the wild type protein"
 XX
 XX WO200198338-A2.
 XX
 XX 27-DEC-2001.
 XX
 XX 18-JUN-2001; 2001WO-EP006848.
 XX
 XX 19-JUN-2000; 2000DK-00000953.
 PR 20-JUN-2000; 2000US-0212681P.
 PR 10-MAY-2001; 2001DK-00000739.
 PR 10-MAY-2001; 2001US-0290170P.
 XX
 XX (BIOI-) BIOIMAGE AS.
 XX
 XX Bjorn SP, Pagliaro L, Thastrup O;
 XX
 DR WPI; 2002-098224/13.
 DR N-PSDB; AAD28162.
 XX
 XX Novel fluorescent protein in in vitro assay for measuring protein kinase
 PT activity or dephosphorylation activity, or for measuring protein
 PT redistribution, has a green fluorescent protein with F64L and E222G
 PT mutation.
 XX
 XX Example 1; Page 35; 41pp; English.
 XX
 XX The invention relates to a fluorescent protein derived from green
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
 CC at F64L and E222G has a bigger compared to other GFP's making it very
 CC suitable for high throughput screening due to better resolution. The
 CC fluorescent protein is useful in invitro assays for measuring protein
 CC kinase activity or dephosphorylation activity, or for measuring protein
 CC redistribution. The fluorescent protein is useful in studying cellular
 CC functions in living cells; as protein tags in transgenic animals, living
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.
 CC The fluorescent protein is also useful as a cell or organelle integrity
 CC marker, a marker for changes in cell morphology, as transfection marker,
 CC and as a marker to be used in combination with fluorescence activated
 CC cell sorting (FACS). The novel proteins can also be used as reporters to
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
 CC protein is also useful as markers in transcriptional and translational
 CC fusions for performing transposon vector mutagenesis and as a reporter
 CC for bacterial detection. Transposons encoding the fluorescent protein are
 CC useful for screening promoters and for tagging plasmids and chromosomes.
 CC The fluorescent protein engineered into the genome of a phage is useful
 CC for designing diagnostic tool. The present sequence is enhanced F64L
 CC Jellyfish green fluorescent protein (GFP) mutant
 XX
 XX Sequence 239 AA;
 SQ
 Query Match 99.0%; Score 1261; DB 5; Length 239;
 Best Local Similarity 99.2%; Pred.No. 3.1e-121;
 Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MWSKGELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGLPVPWPT 60
 DB 1 MWSKGELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGLPVPWPT 60
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNKTRAEVKEGDTL 120
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNKTRAEVKEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGVSQLA 180

DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 QY 181 DHYQQNTPIGDGFVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPIGDGFVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
 RESULT 6
 ID AAB22882
 XX AAB22882 standard; protein; 239 AA.
 AC AAB22882;
 XX
 DT 10-JAN-2001 (first entry)
 XX
 DE Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.
 XX
 KW Bioreporter protein; fusion protein; recognition site;
 KW cellular targeting sequence; cellular localisation; fluorescent protein;
 KW protease activity detection; toxin detection; cellular stress detection;
 KW drug discovery; cell based screening.
 XX
 OS Aequorea victoria.
 OS Synthetic.
 XX
 PN WO200050872-A2.
 XX
 PD 31-AUG-2000.
 XX
 PF 25-FEB-2000; 2000WO-US004794.
 XX
 PR 26-FEB-1999; 99US-0122152P.
 PR 08-MAR-1999; 99US-0123399P.
 PR 12-JUL-1999; 99US-00352171.
 XX
 XX (CELL-) CELLOMICS INC.
 PA
 XX Giuliano KA, Kapur R;
 PI WPI; 2000-594086/56.
 XX N-PSDB; AAA93373.
 DR
 XX Automated cell-based characterization of toxin by contacting cells
 PT containing luminescent reporter molecules with test substance and
 PT analyzing optically.
 XX
 XX Example 11; Fig 29A; 336pp; English.
 CC
 CC The invention relates to systems, methods and reagents for cell-based
 CC screening or detection of compounds which affect particular biological
 CC functions. The methods of the invention utilise fluorescent bioreporter
 CC molecules which, when acted on by a compound of interest, cause an
 CC alteration in the cellular distribution of at least the fluorescent
 CC moiety. In one embodiment, the biosensors comprise heat shock proteins
 CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent
 CC protein (GFP), or derivatives thereof). Such biosensors are located in
 CC the cytoplasm, but on stress activation translocate to the nucleus. In
 CC another embodiment bioreporter proteins can be used to detect protease
 CC activity. Such protease bioreporter fusion proteins comprise one or more
 CC fluorescent proteins; a recognition signal which is cleaved by the
 CC protease; and at least one cellular localisation signal. The latter two
 CC components may be components of a single protein which is acted upon by
 CC the protease, or may be from heterologous sources. Due to the
 CC localisation signal, the bioreporter protein is localised to a particular
 CC region of the cell. Once acted on by the protease of interest, the
 CC fluorescent protein is cleaved from the localisation sequence, and is
 CC free to migrate to other locations within the cell. The presence of a
 CC second localisation signal attached to the fluorescent protein enables
 CC the fluorescent protein to be directed to a different cellular
 CC compartment after cleavage of the protease recognition sequence. The
 CC change in distribution of the fluorescent protein can be detected using
 CC imaging methods with a high degree of spatial resolution. The methods and

CC biosensors of the invention can be used to investigate a wide range of
CC cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased quantity
CC and speed of data collection and faster evaluation of drug candidates.
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used
CC as components of biosensor fusion proteins of the invention
XX
SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTLPKFTTGTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTLPKFTTGTGKLPVWPWT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVLA 180
QY 181 DHYQQTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
DB 181 DHYQQTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 7
AA54349
ID AAY54349 standard; protein; 239 AA.
XX AAY54349;
DT 06-APR-2000 (first entry)
XX Amino acid sequence of the mutant green fluorescent protein EGFP.
DE Fluorescent protein; green fluorescent protein; emission intensity;
KW fluorescence; pH detection; pH sensor; EGFP.
OS Synthetic.
OS Aequorea victoria.

Key Location/Qualifiers
FH Misc-difference 65
FT /note= "wild type Phe substituted with Leu"
FT Misc-difference 66
FT /note= "wild type Ser substituted with Thr"
FT Misc-difference 232
FT /note= "wild type His substituted with Leu"
XX
XX WO964592-A2.
XX
XX 16-DEC-1999.
PD
XX 08-JUN-1999; 99WO-US012850.
XX
XX 09-JUN-1998; 98US-00094359.
XX 13-OCT-1998; 98US-00172063.
XX
XX (REGC) UNIV CALIFORNIA.
XX (UYOR-) UNIV OREGON STATE.

PI Tsien RY, Llopis J, Wachter RM;
XX
XX WPI: 2000-116540/10.
DR N-PSDB; AA245642.
XX
XX New functional engineered green fluorescent proteins, used for measuring
XX the pH in biological samples and cells.
XX
XX Disclosure; Page 9; 89pp; English.
XX
XX The present sequence represents a functional engineered fluorescent
XX protein based on the Aequorea green fluorescent protein (GFP). The
XX emission intensity changes as pH varies between 5 and 10 of the present
XX protein are novel. The functional engineered fluorescent proteins show
XX reversible changes in fluorescence over physiological pH ranges. They can
XX be used for determining the pH of samples and cells. The polynucleotides
XX can also be used to produce transgenic animals. The fluorescent protein
XX pH sensors can be delivered to cells in the form of polynucleotides
XX encoding the protein sensor fused to a targeting signal. The targeting
XX signal directs the expression of the protein sensors to restricted cell
XX locations. This makes it possible to measure the pH of a precisely
XX defined cellular region or organelle
XX
SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTLPKFTTGTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTLPKFTTGTGKLPVWPWT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVLA 180
QY 181 DHYQQTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
DB 181 DHYQQTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 8
AA54349
ID AAY79584 standard; peptide; 239 AA.
XX AAY79584;
XX AAY79584;
DT 29-AUG-2000 (first entry)
XX EGFP signal domain.
DE EGFP signal domain.
XX
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;
XX analysis; drug discovery.
OS Unidentified.
XX
XX WO200026408-A2.
XX
XX 11-MAY-2000.
PD
XX 29-OCT-1999; 99WO-US025431.
XX
XX 30-OCT-1998; 98US-0106308P.
XX 26-MAY-1999; 99US-0136078P.
XX
XX (CELL-) CELLOMICS INC.
XX
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;
PI

XX. WPI; 2000-365644/31.
DR N-PSDB; AAA27573.
XX Recombinant nucleic acid encoding a protease biosensor useful for
PT fluorescence based cell and molecular biochemical assays for drug
PT discovery comprising three operably linked nucleic acid sequences.
XX
XX Claim 14; Fig 29A; 218pp; English.
XX
XX The present sequence is that of the EGFP signal domain, which can be
CC included in novel recombinant protease biosensors (PBs) of the invention.
CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)
CC comprising at least 1 detectable polypeptide signal such as the present
CC sequence; a second domain (see AAY79588-622) comprising at least 1
CC protease recognition site; and a third domain (see AAY79623-37)
CC comprising at least 1 reactant target sequence. A recombinant nucleic
CC acid (see AAA27627-43) encoding the PB, an expression vector, and a
CC genetically engineered host cell are also claimed. A claimed method for
CC identifying compounds that modify protease activity in a cell involves
CC contacting a host cell that possesses the recombinant PB with a test
CC compound, and determining the PB distribution in the host cell, where
CC changes in the distribution of the PB are correlated with modification of
CC protease activity by the test compound. Claimed kits for identifying
CC compounds that modify protease activity in a host cell include the
CC recombinant nucleic acid, or the recombinant PB, or the vector, or the
CC host cell. The PB is useful in high content screens to detect in vivo
CC activation of enzymatic activity, and to identify specific activity based
CC on cleavage of a known recognition motif
XX
XX Sequence 239 AA;
SQ

Query Match 98.7%; Score 1258; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSDPNKRDHVMVLAAGVTAAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSDPNKRDHVMVLAAGVTAAAGITLGMDELYK 239

RESULT 9
AAB50804
ID AAB50804 standard; protein; 239 AA.
XX
XX AAB50804;
AC
XX
DT 14-MAR-2001 (first entry)
XX
DE Jellyfish GFP mutant EGFP.
XX
KW Aequorea victoria; jellyfish; fluorescent protein indicator;
KW green fluorescent protein; GFP; linker moiety; sensor;
KW calmodulin-binding domain; mutant; mutein.
XX
XX Aequorea victoria.
OS
XX WO200071565-A2.
FN
XX 30-NOV-2000.
PD
XX

PF 17-MAY-2000; 2000WO-US013684.
XX
XX 21-MAY-1999; 99US-00316919.
PR 21-MAY-1999; 99US-00316920.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Tsien RY, Baird GA;
PI
XX WPI; 2001-032017/04.
DR N-PSDB; AAC90488.
XX
XX Novel fluorescent proteins comprising a sensor protein inserted into
PT them, useful for measuring the response of a sensor biological, chemical,
PT electrical or physiological parameter in vivo or in vitro.
XX
XX Disclosure; Page 24; 94pp; English.
XX
XX The present sequence is a fluorescent protein used in the construction of
CC a fluorescent protein indicator. The indicator comprises a sensor
CC polypeptide that is responsive to a chemical, biological, electrical or
CC physiological parameter, and a fluorescence protein functional group. The
CC sensor polypeptide is operatively inserted into the fluorescent moiety.
CC The fluorescent indicator is useful for detecting the presence of a
CC response inducing member in a sample. The method involves contacting the
CC sample with the indicator and detecting a change in fluorescence, in
CC which a change is indicative of the effect of the parameter on the sensor
CC polypeptide. The novel fluorescent proteins are advantageous due to their
CC reduced size as compared to the FRET (fluorescence resonance energy
CC transfer)-based sensors
XX
XX Sequence 239 AA;
SQ

Query Match 98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSDPNKRDHVMVLAAGVTAAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQSDPNKRDHVMVLAAGVTAAAGITLGMDELYK 239

RESULT 10
AAB85900
ID AAB85900 standard; protein; 239 AA.
XX
XX AAB85900;
AC
XX
DT 30-NOV-2001 (first entry)
XX
DE A. victoria green fluorescent protein (GFP) and linker sequence.
XX
KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;
KW green fluorescent protein.
XX
XX Synthetic.
OS
XX Aequorea victoria.
FN
XX WO200168706-A1.
PD
XX

PD 20-SEP-2001.
 XX
 PF 14-MAR-2001; 2001WO-US008071.
 XX
 PR 15-MAR-2000; 2000US-0189698P.
 XX
 PA (MERI) MERCK & CO INC.
 XX
 XX Marsh DJ;
 XX
 XX WPI; 2001-565791/63.
 DR N-PSDB; AAH47304.
 XX
 XX Fusion proteins comprising melanin concentrating hormone receptor
 PT peptides and fluorescent proteins, useful for identifying appetite
 PT stimulants.
 XX
 PS Claim 2; Page 14; 71pp; English.
 XX
 CC The invention provides melanin concentrating hormone (MCH) receptor
 CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
 CC MCHR polypeptide regions from different species. The MCHR fusion protein
 CC comprise MCHR polypeptide region and a fluorescent polypeptide region
 CC joined directly, or via a linker, to the carboxy side of the MCHR
 CC polypeptide region. The MCHR fusion proteins can be expressed by standard
 CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
 CC regulation of MCH activity stimulates food intake. The present sequence
 CC represents a A. victoria green fluorescent protein (GFP) and a linker
 CC sequence
 XX
 XX Sequence 239 AA;
 SQ
 Query Match 98.7%; Score 1258; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 6.3e-121; Indels 0; Gaps 0;
 Matches 236; Conservative 1; Mismatches 2;
 QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239
 DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITLGMDELYK 239
 RESULT 11
 AAB31171
 ID AAB31171 standard; protein; 239 AA.
 XX
 AC AAB31171;
 XX
 XX 02-APR-2001 (first entry)
 DT
 XX Amino acid sequence of a green fluorescent protein (GFP).
 DE
 XX Growth rate; death rate; reporter gene; luminescent protein;
 KW fluorescent product; luciferase; green fluorescent protein; GFP.
 XX
 XX Aequorea victoria.
 OS
 XX WO200075367-A1.
 PN
 XX 14-DEC-2000.
 PD
 XX 07-JUN-2000; 2000WO-FI000507.
 PF

XX 07-JUN-1999; 99FI-00001296.
 PR (LILI/) LILIUS E.
 PA (VIRT/) VIRT M.
 XX
 PI Lilius E, Virta M;
 XX
 XX WPI; 2001-061737/07.
 DR N-PSDB; AAC86954.
 XX
 XX Assessing growth and death rates of a micro-organism in a desired
 PT environment, by introducing 2 reporter genes encoding luminescent and
 PT fluorescent products and detecting luminescent fluorescence.
 XX
 PS Disclosure; Page 27; 32pp; English.
 XX
 CC The specification describes a method for assessing the growth rate and
 CC death rate of a micro-organism within a predetermined time period in a
 CC desired environment. The method comprises introducing at least two
 CC reporter genes encoding luminescent and/or fluorescent products into the
 CC micro-organisms, incubating the micro-organism within the desired
 CC environment, and detecting luminescence and/or fluorescence after a
 CC predetermined time period. Use of two different markers within a micro-
 CC organism enables the differentiation between growth and death rates. The
 CC method is used to assess the growth rate and death rate of a micro-
 CC organism within a predetermined time period in a desired environment. The
 CC present sequence represents a green fluorescent protein (GFP), and is
 CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
 CC and is used in the method of the invention
 XX
 XX Sequence 239 AA;
 SQ
 Query Match 98.7%; Score 1258; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 6.3e-121; Indels 0; Gaps 0;
 Matches 236; Conservative 1; Mismatches 2;
 QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
 QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239
 DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITLGMDELYK 239
 RESULT 12
 AAG66198
 ID AAG66198 standard; protein; 239 AA.
 XX
 AC AAG66198;
 XX
 XX 17-JUN-2002 (first entry)
 DT
 XX A. victoria green fluorescent protein (EGFP).
 DE
 XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
 KW green fluorescent protein; EGFP.
 XX
 XX Aequorea victoria.
 OS
 XX JP20002045189-A.
 PN
 XX 12-FEB-2002.
 PD
 XX

XX FH Key Location/Qualifiers
FT Misc-difference 1. .3
FT /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"
FT
FT Misc-difference 65
FT /note= "GFP Phe64 is replaced by Leu"
FT Misc-difference 66
FT /note= "GFP Ser65 is replaced by Thr"
XX
PN EP1178109-A1.
XX
XX 06-FEB-2002.
XX
XX 03-AUG-2001; 2001EP-00306650.
XX
XX 04-AUG-2000; 2000JP-00237166.
XX
XX (RIKE) RIKEN KK.
XX
XX Miyawaki A, Sawano A;
XX
XX WPI; 2002-208112/27.
XX N-PSDB; AAD27910.
XX
XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.
XX
XX Example 1; Page 13-14; 31pp; English.
XX
XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments, and a double-stranded of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is CC Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria
XX
SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKGELFTGVVPIVLVDGVDVNGHKFSVSGEGEDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKGELFTGVVPIVLVDGVDVNGHKFSVSGEGEDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTLYGVCFSRYPDHMKQHDFFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLGFTVTAAGTTLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLGFTVTAAGTTLGMDELYK 239

QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLGFTVTAAGTTLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLGFTVTAAGTTLGMDELYK 239

RESULT 15
AAE34958
ID AAE34958 standard; protein; 239 AA.
XX
AC AAE34958;
XX
XX 28-MAY-2003 (first entry)
DE Aequorea victoria enhanced green fluorescent protein (EGFP).
XX
XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;
KW Kinase; enhanced green fluorescent protein; EGFP.
XX
OS Aequorea victoria.
XX
XX WO200295058-A2.
XX
XX 28-NOV-2002.
XX
XX 24-MAY-2002; 2002WO-US016955.
XX
XX 24-MAY-2001; 2001US-00865291.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Tsien RV, Ting AV, Zhang J;
XX
XX WPI; 2003-148474/14.
DR N-PSDB; AAD53428.
XX
XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.
XX
XX Disclosure; Col 56-57; 38pp; English.
XX
XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or in operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention
XX
SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 6; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKGELFTGVVPIVLVDGVDVNGHKFSVSGEGEDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKGELFTGVVPIVLVDGVDVNGHKFSVSGEGEDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTLYGVCFSRYPDHMKQHDFFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLGFTVTAAGTTLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLGFTVTAAGTTLGMDELYK 239

Db 181 DHYQQNTPIGDGPVLLFPDNHYLSTQSALSXDPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:13
Job time : 48.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-222a
Perfect score: 1274
Sequence: 1 MVSKEELFTGVVPIVELD.....VLAFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/FCUTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1258	98.7	239	3	US-09-172-063-3 Sequence 3, Appli
2	1258	98.7	239	4	US-09-513-783A-46 Sequence 46, Appl
3	1258	98.7	239	4	US-09-316-919-4 Sequence 4, Appli
4	1258	98.7	239	4	US-09-602-641-3 Sequence 3, Appli
5	1258	98.7	239	4	US-09-920-922-2 Sequence 2, Appli
6	1258	98.7	281	3	US-09-062-102-1 Sequence 1, Appli
7	1258	98.7	281	4	US-09-364-946-1 Sequence 1, Appli
8	1258	98.7	294	4	US-09-513-783A-2 Sequence 2, Appli
9	1258	98.7	323	3	US-09-172-063-21 Sequence 21, Appl
10	1258	98.7	323	4	US-09-602-641-21 Sequence 21, Appl
11	1258	98.7	364	3	US-09-085-305-6 Sequence 6, Appli
12	1258	98.7	379	4	US-09-417-197-129 Sequence 129, Appl
13	1258	98.7	434	4	US-09-800-170-48 Sequence 129, Appl
14	1258	98.7	442	4	US-09-417-197-127 Sequence 127, App
15	1258	98.7	459	4	US-09-513-783A-170 Sequence 170, App
16	1258	98.7	544	4	US-09-417-197-113 Sequence 113, App
17	1258	98.7	544	4	US-09-417-197-115 Sequence 115, App
18	1258	98.7	604	4	US-09-417-197-59 Sequence 59, Appl
19	1258	98.7	605	4	US-09-417-197-41 Sequence 41, Appl
20	1258	98.7	606	4	US-09-417-197-65 Sequence 65, Appl
21	1258	98.7	607	4	US-09-417-197-47 Sequence 47, Appl
22	1258	98.7	630	4	US-09-417-197-63 Sequence 63, Appl
23	1258	98.7	631	4	US-09-417-197-39 Sequence 39, Appl
24	1258	98.7	633	4	US-09-417-197-45 Sequence 45, Appl
25	1258	98.7	635	4	US-09-417-197-125 Sequence 125, App
26	1258	98.7	642	2	US-08-818-253-2 Sequence 2, Appli
27	1258	98.7	642	2	US-08-818-253-6 Sequence 6, Appli

28	1258	98.7	642	3	US-08-818-253-2	Sequence 2, Appli
29	1258	98.7	642	3	US-08-818-253-6	Sequence 6, Appli
30	1258	98.7	652	2	US-08-818-253-4	Sequence 4, Appli
31	1258	98.7	652	3	US-08-818-252-4	Sequence 4, Appli
32	1258	98.7	718	4	US-09-417-197-75	Sequence 75, Appl
33	1258	98.7	719	4	US-09-417-197-51	Sequence 51, Appl
34	1258	98.7	726	4	US-09-417-197-71	Sequence 71, Appl
35	1258	98.7	727	4	US-09-417-197-139	Sequence 139, App
36	1258	98.7	783	4	US-09-513-783A-176	Sequence 176, App
37	1258	98.7	797	4	US-09-417-197-141	Sequence 141, App
38	1258	98.7	797	4	US-09-417-197-143	Sequence 143, App
39	1258	98.7	798	4	US-09-417-197-77	Sequence 77, Appl
40	1258	98.7	805	4	US-09-513-783A-178	Sequence 178, App
41	1258	98.7	806	4	US-09-417-197-51	Sequence 53, Appl
42	1258	98.7	836	4	US-09-417-197-61	Sequence 61, Appl
43	1258	98.7	842	4	US-09-417-197-43	Sequence 43, Appl
44	1258	98.7	843	4	US-09-417-197-117	Sequence 117, App
45	1258	98.7	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1
US-09-172-063-3
; Sequence 3, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Lloplis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-172-063-3

Query Match	98.7%	Score	1258	DB	3	Length	239
Best Local Similarity	98.7%	Pred. No.	4e-128				
Matches	236	Conservative	1	Mismatches	2	Indels	0
							Gaps 0;
QY	1	MVSKGEELFTGVVPIVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVWPWT	60				
Db	1	MVSKGEELFTGVVPIVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVWPWT	60				
QY	61	LVTTLTSGVQCPSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL	120				
Db	61	LVTTLTSGVQCPSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL	120				
QY	121	VNRIELKIDFKEDGNILGHKLEYNHNHVYIMADKQKNGIKVNFKIRHNIEDSGVQLA	180				
Db	121	VNRIELKIDFKEDGNILGHKLEYNHNHVYIMADKQKNGIKVNFKIRHNIEDSGVQLA	180				
QY	181	DHYQONTPTIGDGPVLLPNHNYLSTQSALSKDNEKRDHMLAGFVTAAGITLGMDELYK	239				
Db	181	DHYQONTPTIGDGPVLLPNHNYLSTQSALSKDNEKRDHMLAGFVTAAGITLGMDELYK	239				

```

RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 4e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLGAFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLGAFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 4e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLGAFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLGAFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 4e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLGAFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLGAFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9

```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-920-922-2

Query Match      98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 4e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
DB 61 LVTTLTIGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
QY 121 VNRLEKLGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKLGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangliang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match      98.7%; Score 1258; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 5.1e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
DB 61 LVTTLTIGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
QY 121 VNRLEKLGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKLGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangliang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match      98.7%; Score 1258; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 5.1e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
DB 61 LVTTLTIGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
QY 121 VNRLEKLGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKLGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kaput, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match      98.7%; Score 1258; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 5.5e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
```

Db 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239

RESULT 9

US-09-172-063-21

; Sequence 21, Application US/09172063

; Patent No. 6150176

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.

; APPLICANT: Remington, S. James

; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR

; FILE REFERENCE: 07257/071001

; CURRENT APPLICATION NUMBER: US/09/172,063

; CURRENT FILING DATE: 1998-10-13

; EARLIER APPLICATION NUMBER: 09/094,359

; EARLIER FILING DATE: 1998-06-09

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 21

; LENGTH: 323

; TYPE: PRT

; ORGANISM: Aequorea victoria

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (0)...(0)

; OTHER INFORMATION: GT-EGFP

US-09-172-063-21

Query Match 98.7%; Score 1258; DB 3; Length 323;

Best Local Similarity 98.7%; Pred. No. 6.3e-128;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 144
QY 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204
QY 121 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21

; Sequence 21, Application US/09602641

; Patent No. 6608189

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (0)...(0)

; OTHER INFORMATION: GT-EGFP

US-09-602-641-21

Query Match 98.7%; Score 1258; DB 4; Length 323;

Best Local Similarity 98.7%; Pred. No. 6.3e-128;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 144
QY 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204
QY 121 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6

; Sequence 6, Application US/09085305

; Patent No. 6191269

; GENERAL INFORMATION:

; APPLICANT: Pollock, Allan

; APPLICANT: Lovett, David H.

; APPLICANT: Turck, Johanna

; TITLE OF INVENTION: Selective Induction of Apoptosis in

; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal

; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bozicevic & Reed, LLP

; STREET: 285 Hamilton Ave, Suite 200

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/085,305

; FILING DATE: 29-MAY-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Francis, Carol L
REGISTRATION NUMBER: 36,513
REFERENCE/DOCKET NUMBER: 6510/102US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-085-305-6

Query Match 98.7%; Score 1258; DB 3; Length 364;
Best Local Similarity 98.7%; Pred. NO. 7.6e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60

DB 126 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 185

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

DB 186 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 180

DB 246 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 305

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQALSQSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 239

DB 306 DHYQONTPIGDGPVLLPDNHYLSTQALSQSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 364

RESULT 12

US-09-417-197-129

Sequence 129, Application US/09417197

Patent No. 6518021

GENERAL INFORMATION:

APPLICANT: Ole THASTRUP, et al.

TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An

TITLE OF INVENTION: On A Cellular Response

FILE REFERENCE: 3759-0110P

CURRENT FILING DATE: US/09/417,197

NUMBER OF SEQ ID NOS: 143

SOFTWARE: PatentIn version 3.0

SEQ ID NO 129

LENGTH: 379

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: actin-binding-domain-EGFP fusion

US-09-417-197-129

Query Match 98.7%; Score 1258; DB 4; Length 379;
Best Local Similarity 98.7%; Pred. NO. 8e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60

DB 141 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 200

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

DB 201 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 180

DB 261 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 320

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQALSQSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 239

DB 321 DHYQONTPIGDGPVLLPDNHYLSTQALSQSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 379

RESULT 13

US-09-800-170-48

Sequence 48, Application US/09800170

Patent No. 6481667

GENERAL INFORMATION:

APPLICANT: Kinsella, Todd

TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES

FILE REFERENCE: A-68614-1/DJB/RMS/RMK

CURRENT APPLICATION NUMBER: US/09/800,170

CURRENT FILING DATE: 2001-08-28

PRIOR APPLICATION NUMBER: US 60/187,130

PRIOR FILING DATE: 2000-03-06

NUMBER OF SEQ ID NOS: 90

SOFTWARE: PatentIn version 3.1

SEQ ID NO 48

LENGTH: 434

TYPE: PRT

ORGANISM: Synechocystis PCC6803

US-09-800-170-48

Query Match 98.7%; Score 1258; DB 4; Length 434;

Best Local Similarity 98.7%; Pred. No. 9.9e-128;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60

DB 196 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 255

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

DB 256 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 180

DB 316 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 375

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQALSQSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 239

DB 376 DHYQONTPIGDGPVLLPDNHYLSTQALSQSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 434

RESULT 14

US-09-417-197-127

Sequence 127, Application US/09417197

Patent No. 6518021

GENERAL INFORMATION:

APPLICANT: Ole THASTRUP, et al.

TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An

TITLE OF INVENTION: On A Cellular Response

FILE REFERENCE: 3759-0110P

CURRENT APPLICATION NUMBER: US/09/417,197

CURRENT FILING DATE: 1999-10-07

NUMBER OF SEQ ID NOS: 143

SOFTWARE: PatentIn version 3.0

SEQ ID NO 127

LENGTH: 442

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: EGFP-RhoA fusion

US-09-417-197-127

Query Match 98.7%; Score 1258; DB 4; Length 442;

Best Local Similarity 98.7%; Pred. No. 1e-127;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60

```
Db      1  MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVPWPT 60
QY      61  LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db      61  LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY    181  DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAAGVTAAGITLGMDELYK 239
Db     181  DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
```

RESULT 15

```
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170
```

```
Query Match      98.7%; Score 1258; DB 4; Length 459;
Best Local Similarity 98.7%; Pred.No. 1.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1  MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVPWPT 60
Db      1  MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVPWPT 60
QY     61  LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db     61  LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY    131  DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAAGVTAAGITLGMDELYK 239
Db    131  DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
```

Search completed: June 21, 2004, 16:04:05
Job time : 13.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds
(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-222a

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVPIVLVELD.....VLaGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1269	99.6	239	9	US-09-887-784-4
2	1269	99.6	239	12	US-10-296-953-4
3	1269	99.6	363	14	US-10-270-223-6
4	1269	99.6	893	14	US-10-257-909A-30
5	1269	99.6	1132	14	US-10-257-909A-32
6	1261	99.0	239	9	US-09-887-784-2
7	1261	99.0	239	12	US-10-296-953-2
8	1258	98.7	239	9	US-09-920-922-2
9	1258	98.7	239	9	US-09-999-745-4
10	1258	98.7	239	10	US-09-866-538-4
11	1258	98.7	239	10	US-09-797-496B-2
12	1258	98.7	239	10	US-09-794-308-4
13	1258	98.7	239	10	US-09-865-291-4
14	1258	98.7	239	12	US-10-457-982-3
15	1258	98.7	239	14	US-10-121-258-13

16	1258	98.7	239	14	US-10-221-461-7	Sequence 7, Appli
17	1258	98.7	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1258	98.7	239	14	US-10-177-390-2	Sequence 2, Appli
19	1258	98.7	239	14	US-10-338-411-3	Sequence 3, Appli
20	1258	98.7	239	15	US-10-370-570-4	Sequence 4, Appli
21	1258	98.7	239	15	US-10-389-640-3	Sequence 3, Appli
22	1258	98.7	259	14	US-10-314-861-11	Sequence 11, Appli
23	1258	98.7	281	12	US-09-931-232-1	Sequence 37, Appli
24	1258	98.7	288	14	US-10-314-861-37	Sequence 11, Appli
25	1258	98.7	293	14	US-10-314-861-35	Sequence 35, Appli
26	1258	98.7	294	14	US-10-100-957A-2	Sequence 35, Appli
27	1258	98.7	295	14	US-10-314-861-39	Sequence 2, Appli
28	1258	98.7	299	14	US-10-314-861-33	Sequence 39, Appli
29	1258	98.7	305	14	US-10-314-861-31	Sequence 33, Appli
30	1258	98.7	308	14	US-10-033-717-35	Sequence 35, Appli
31	1258	98.7	311	14	US-10-314-861-29	Sequence 29, Appli
32	1258	98.7	320	14	US-10-338-411-11	Sequence 11, Appli
33	1258	98.7	320	15	US-10-389-640-11	Sequence 11, Appli
34	1258	98.7	323	12	US-10-457-982-21	Sequence 21, Appli
35	1258	98.7	323	14	US-10-338-411-7	Sequence 7, Appli
36	1258	98.7	323	14	US-10-338-411-13	Sequence 13, Appli
37	1258	98.7	323	15	US-10-389-640-7	Sequence 7, Appli
38	1258	98.7	323	15	US-10-389-640-13	Sequence 13, Appli
39	1258	98.7	324	14	US-10-314-861-16	Sequence 16, Appli
40	1258	98.7	345	14	US-10-338-411-5	Sequence 5, Appli
41	1258	98.7	345	15	US-10-389-640-5	Sequence 5, Appli
42	1258	98.7	346	14	US-10-338-411-9	Sequence 9, Appli
43	1258	98.7	346	15	US-10-389-640-9	Sequence 9, Appli
44	1258	98.7	359	14	US-10-033-717-33	Sequence 33, Appli
45	1258	98.7	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequoria Victoria
US-09-887-784-4

Query Match 99.6%; Score 1269; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.6e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKEELFTGVVPIVLVELDGVNKGKFSVSGEGDATYGKLTCLKFICTTGLKLPVPWPT	60
Db	1	MVSKEELFTGVVPIVLVELDGVNKGKFSVSGEGDATYGKLTCLKFICTTGLKLPVPWPT	60
Qy	61	LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVYKFEQDITL	120
Db	61	LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVYKFEQDITL	120
Qy	121	VNRIELKIDFDKEDGNILGHKLEYNNSVIMADKQKNGIKVNPKIRHNIEDGSVQLA	180
Db	121	VNRIELKIDFDKEDGNILGHKLEYNNSVIMADKQKNGIKVNPKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPIGDGFPVLLPDNHYLSTQSALSQDNREKRDHVLGAFVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDGFPVLLPDNHYLSTQSALSQDNREKRDHVLGAFVTAAGITLGMDELYK	239

```
RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match          99.6%; Score 1269; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.6e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 3
US-10-270-223-6
; Sequence 5, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: Bioimage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACTING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match          99.6%; Score 1269; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 4.7e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.6%; Score 1269; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 1.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
```

```
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.6%; Score 1269; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 1.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
```



```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.6%; Score 1269; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 2.4e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 953

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHHYSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJOERN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          99.0%; Score 1261; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.0%; Score 1261; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.7%; Score 1258; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.6e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
```

```
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.0%; Score 1261; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.7%; Score 1258; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.6e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
```


APPLICANT: TSIEH, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match 98.7%; Score 1258; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.6e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLTKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLTKFKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
Sequence 4, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEH, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REGEN1550
CURRENT APPLICATION NUMBER: US/09/865,291
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match 98.7%; Score 1258; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.6e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLTKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLTKFKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239
RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US20030212365A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Atsushi
APPLICANT: Llopis, Juan
APPLICANT: Wachter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR APPLICATION NUMBER: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 98.7%; Score 1258; DB 12; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.6e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLTKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLTKFKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 15
US-10-121-258-13
Sequence 13, Application US/10121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CPI
CURRENT APPLICATION NUMBER: US/10/121,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.7%; Score 1258; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. NO. 3.6e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGKLTTLKFICTTGKLPVPWPT 60
Db      1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGKLTTLKFICTTGKLPVPWPT 60

QY      61 LVTTLSYGVOCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNVKTAEVKFEGDTL 120
Db      61 LVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNVKTAEVKFEGDTL 120

QY      121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db      121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
Db      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLEFVTAAGITLGMDELYK 239
```

Search completed: June 21, 2004, 16:09:30
Job time : 35.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds
(without alignments)
2224.817 Million cell updates/sec

Title: US-09-887-784-222A

Perfect score: 1274

Sequence: 1 MVSKEBELFTGVVILVELD.....VLagFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1236	97.0	238	1 JQ1514	green-fluorescent
2	1106	8.3	785	2 H72228	hypothetical prote
3	91.5	7.2	861	2 H64102	leucine-tRNA ligas
4	90	7.1	632	2 T06586	DNA-binding protei
5	89.5	7.0	887	2 E82590	leucyl-tRNA synthe
6	89	7.0	578	1 I40794	dihydrolipoamide d
7	88.5	6.9	655	2 D83917	DNA topoisomerase
8	87.5	6.9	370	2 E70390	iron-sulfur cofact
9	87.5	6.9	860	2 AC0382	leucyl-tRNA synthe
10	87.5	6.9	2222	1 A36028	DNA-directed DNA p
11	87.5	6.9	2573	2 D71614	hypothetical prote
12	86	6.8	357	2 G81355	tRNA (uracil-5)-m
13	86	6.8	461	2 T06936	photosystem II chl
14	85.5	6.7	788	1 JQV1HH	DNA-directed DNA p
15	85.5	6.7	797	2 J4078	protective surface
16	85.5	6.7	808	2 F64102	inter-alpha-trypsi
17	85.5	6.7	889	2 JC5576	hypothetical prote
18	85	6.7	281	2 AD2052	synthetohymenotropi
19	84.5	6.6	310	2 S68225	hypothetical prote
20	84.5	6.6	435	2 C97354	oligoendopeptidase
21	84.5	6.6	613	2 A99552	cellulase (EC 3.2.
22	84.5	6.6	941	2 S29043	imidazoleglycerol-
23	84	6.6	353	2 J84941	S-layer protein pr
24	84	6.6	874	2 J49330	water-stress-induc
25	83.5	6.6	263	2 S53488	nitrogenase (EC 1.
26	83.5	6.6	534	1 NICLMA	DNA-directed DNA p
27	83.5	6.6	836	1 JQV1LD	IgA Fc receptor pr
28	83.5	6.6	1134	2 A60234	IgA Fc receptor pr
29	83.5	6.6	1164	1 FCSONG	IgA Fc receptor pr

hypothetical prote
hypothetical prote
hypothetical prote
lipoxigenase (SC 1
probable lipoxigen
neuraminidase, pro
hypothetical prote
hypothetical prote
photosystem II chl
hypothetical prote
carboxy-terminal p
proprotein convert
conserved hypothet
hypothetical prote
probable exported

ALIGNMENTS

RESULT 1

JQ1514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C;Species: Aequorea victoria
C;Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C;Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331
R;Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A;Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A;Reference number: JQ1514; MUID:92175527; PMID:1347277
A;Accession: JS0692
A;Molecule type: DNA
A;Residues: 1-107, 'S', 109-238 <PRA1>
A;Cross-references: GB:M62654; NID:g155662; PIDN:AAA27722.1; PID:g155663
A;Accession: JQ1514
A;Molecule type: mRNA
A;Residues: 1-99, 'P', '101-140, 'L', '142-218, 'V', '220-238 <PRA2>
A;Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661
A;Accession: PQ0335
A;Molecule type: protein
A;Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>
R;Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A;Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A;Reference number: S48693; MUID:94364470; PMID:8082767
A;Accession: S48693
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-24, 'Q', '26-156, 'P', '158-171, 'K', '173-238 <INO>
A;Cross-references: GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:g606384
R;Watkins, J.N.; Campbell, A.K.
submitted to the EMBL data Library, January 1995
A;Reference number: S51330
A;Accession: S51330
A;Molecule type: mRNA
A;Residues: 1-13, 'V', '15-24, 'Q', '26-44, 'N', '46-153, 'G', '155-156, 'P', '158-171, 'K', '173-227, 'R', '173-228, 'Q', '173-229, 'R', '173-230, 'Q', '173-231, 'R', '173-232, 'Q', '173-233, 'R', '173-234, 'Q', '173-235, 'R', '173-236, 'Q', '173-237, 'R', '173-238, 'Q', '173-239, 'R', '173-240, 'Q', '173-241, 'R', '173-242, 'Q', '173-243, 'R', '173-244, 'Q', '173-245, 'R', '173-246, 'Q', '173-247, 'R', '173-248, 'Q', '173-249, 'R', '173-250, 'Q', '173-251, 'R', '173-252, 'Q', '173-253, 'R', '173-254, 'Q', '173-255, 'R', '173-256, 'Q', '173-257, 'R', '173-258, 'Q', '173-259, 'R', '173-260, 'Q', '173-261, 'R', '173-262, 'Q', '173-263, 'R', '173-264, 'Q', '173-265, 'R', '173-266, 'Q', '173-267, 'R', '173-268, 'Q', '173-269, 'R', '173-270, 'Q', '173-271, 'R', '173-272, 'Q', '173-273, 'R', '173-274, 'Q', '173-275, 'R', '173-276, 'Q', '173-277, 'R', '173-278, 'Q', '173-279, 'R', '173-280, 'Q', '173-281, 'R', '173-282, 'Q', '173-283, 'R', '173-284, 'Q', '173-285, 'R', '173-286, 'Q', '173-287, 'R', '173-288, 'Q', '173-289, 'R', '173-290, 'Q', '173-291, 'R', '173-292, 'Q', '173-293, 'R', '173-294, 'Q', '173-295, 'R', '173-296, 'Q', '173-297, 'R', '173-298, 'Q', '173-299, 'R', '173-300, 'Q', '173-301, 'R', '173-302, 'Q', '173-303, 'R', '173-304, 'Q', '173-305, 'R', '173-306, 'Q', '173-307, 'R', '173-308, 'Q', '173-309, 'R', '173-310, 'Q', '173-311, 'R', '173-312, 'Q', '173-313, 'R', '173-314, 'Q', '173-315, 'R', '173-316, 'Q', '173-317, 'R', '173-318, 'Q', '173-319, 'R', '173-320, 'Q', '173-321, 'R', '173-322, 'Q', '173-323, 'R', '173-324, 'Q', '173-325, 'R', '173-326, 'Q', '173-327, 'R', '173-328, 'Q', '173-329, 'R', '173-330, 'Q', '173-331, 'R', '173-332, 'Q', '173-333, 'R', '173-334, 'Q', '173-335, 'R', '173-336, 'Q', '173-337, 'R', '173-338, 'Q', '173-339, 'R', '173-340, 'Q', '173-341, 'R', '173-342, 'Q', '173-343, 'R', '173-344, 'Q', '173-345, 'R', '173-346, 'Q', '173-347, 'R', '173-348, 'Q', '173-349, 'R', '173-350, 'Q', '173-351, 'R', '173-352, 'Q', '173-353, 'R', '173-354, 'Q', '173-355, 'R', '173-356, 'Q', '173-357, 'R', '173-358, 'Q', '173-359, 'R', '173-360, 'Q', '173-361, 'R', '173-362, 'Q', '173-363, 'R', '173-364, 'Q', '173-365, 'R', '173-366, 'Q', '173-367, 'R', '173-368, 'Q', '173-369, 'R', '173-370, 'Q', '173-371, 'R', '173-372, 'Q', '173-373, 'R', '173-374, 'Q', '173-375, 'R', '173-376, 'Q', '173-377, 'R', '173-378, 'Q', '173-379, 'R', '173-380, 'Q', '173-381, 'R', '173-382, 'Q', '173-383, 'R', '173-384, 'Q', '173-385, 'R', '173-386, 'Q', '173-387, 'R', '173-388, 'Q', '173-389, 'R', '173-390, 'Q', '173-391, 'R', '173-392, 'Q', '173-393, 'R', '173-394, 'Q', '173-395, 'R', '173-396, 'Q', '173-397, 'R', '173-398, 'Q', '173-399, 'R', '173-400, 'Q', '173-401, 'R', '173-402, 'Q', '173-403, 'R', '173-404, 'Q', '173-405, 'R', '173-406, 'Q', '173-407, 'R', '173-408, 'Q', '173-409, 'R', '173-410, 'Q', '173-411, 'R', '173-412, 'Q', '173-413, 'R', '173-414, 'Q', '173-415, 'R', '173-416, 'Q', '173-417, 'R', '173-418, 'Q', '173-419, 'R', '173-420, 'Q', '173-421, 'R', '173-422, 'Q', '173-423, 'R', '173-424, 'Q', '173-425, 'R', '173-426, 'Q', '173-427, 'R', '173-428, 'Q', '173-429, 'R', '173-430, 'Q', '173-431, 'R', '173-432, 'Q', '173-433, 'R', '173-434, 'Q', '173-435, 'R', '173-436, 'Q', '173-437, 'R', '173-438, 'Q', '173-439, 'R', '173-440, 'Q', '173-441, 'R', '173-442, 'Q', '173-443, 'R', '173-444, 'Q', '173-445, 'R', '173-446, 'Q', '173-447, 'R', '173-448, 'Q', '173-449, 'R', '173-450, 'Q', '173-451, 'R', '173-452, 'Q', '173-453, 'R', '173-454, 'Q', '173-455, 'R', '173-456, 'Q', '173-457, 'R', '173-458, 'Q', '173-459, 'R', '173-460, 'Q', '173-461, 'R', '173-462, 'Q', '173-463, 'R', '173-464, 'Q', '173-465, 'R', '173-466, 'Q', '173-467, 'R', '173-468, 'Q', '173-469, 'R', '173-470, 'Q', '173-471, 'R', '173-472, 'Q', '173-473, 'R', '173-474, 'Q', '173-475, 'R', '173-476, 'Q', '173-477, 'R', '173-478, 'Q', '173-479, 'R', '173-480, 'Q', '173-481, 'R', '173-482, 'Q', '173-483, 'R', '173-484, 'Q', '173-485, 'R', '173-486, 'Q', '173-487, 'R', '173-488, 'Q', '173-489, 'R', '173-490, 'Q', '173-491, 'R', '173-492, 'Q', '173-493, 'R', '173-494, 'Q', '173-495, 'R', '173-496, 'Q', '173-497, 'R', '173-498, 'Q', '173-499, 'R', '173-500, 'Q', '173-501, 'R', '173-502, 'Q', '173-503, 'R', '173-504, 'Q', '173-505, 'R', '173-506, 'Q', '173-507, 'R', '173-508, 'Q', '173-509, 'R', '173-510, 'Q', '173-511, 'R', '173-512, 'Q', '173-513, 'R', '173-514, 'Q', '173-515, 'R', '173-516, 'Q', '173-517, 'R', '173-518, 'Q', '173-519, 'R', '173-520, 'Q', '173-521, 'R', '173-522, 'Q', '173-523, 'R', '173-524, 'Q', '173-525, 'R', '173-526, 'Q', '173-527, 'R', '173-528, 'Q', '173-529, 'R', '173-530, 'Q', '173-531, 'R', '173-532, 'Q', '173-533, 'R', '173-534, 'Q', '173-535, 'R', '173-536, 'Q', '173-537, 'R', '173-538, 'Q', '173-539, 'R', '173-540, 'Q', '173-541, 'R', '173-542, 'Q', '173-543, 'R', '173-544, 'Q', '173-545, 'R', '173-546, 'Q', '173-547, 'R', '173-548, 'Q', '173-549, 'R', '173-550, 'Q', '173-551, 'R', '173-552, 'Q', '173-553, 'R', '173-554, 'Q', '173-555, 'R', '173-556, 'Q', '173-557, 'R', '173-558, 'Q', '173-559, 'R', '173-560, 'Q', '173-561, 'R', '173-562, 'Q', '173-563, 'R', '173-564, 'Q', '173-565, 'R', '173-566, 'Q', '173-567, 'R', '173-568, 'Q', '173-569, 'R', '173-570, 'Q', '173-571, 'R', '173-572, 'Q', '173-573, 'R', '173-574, 'Q', '173-575, 'R', '173-576, 'Q', '173-577, 'R', '173-578, 'Q', '173-579, 'R', '173-580, 'Q', '173-581, 'R', '173-582, 'Q', '173-583, 'R', '173-584, 'Q', '173-585, 'R', '173-586, 'Q', '173-587, 'R', '173-588, 'Q', '173-589, 'R', '173-590, 'Q', '173-591, 'R', '173-592, 'Q', '173-593, 'R', '173-594, 'Q', '173-595, 'R', '173-596, 'Q', '173-597, 'R', '173-598, 'Q', '173-599, 'R', '173-600, 'Q', '173-601, 'R', '173-602, 'Q', '173-603, 'R', '173-604, 'Q', '173-605, 'R', '173-606, 'Q', '173-607, 'R', '173-608, 'Q', '173-609, 'R', '173-610, 'Q', '173-611, 'R', '173-612, 'Q', '173-613, 'R', '173-614, 'Q', '173-615, 'R', '173-616, 'Q', '173-617, 'R', '173-618, 'Q', '173-619, 'R', '173-620, 'Q', '173-621, 'R', '173-622, 'Q', '173-623, 'R', '173-624, 'Q', '173-625, 'R', '173-626, 'Q', '173-627, 'R', '173-628, 'Q', '173-629, 'R', '173-630, 'Q', '173-631, 'R', '173-632, 'Q', '173-633, 'R', '173-634, 'Q', '173-635, 'R', '173-636, 'Q', '173-637, 'R', '173-638, 'Q', '173-639, 'R', '173-640, 'Q', '173-641, 'R', '173-642, 'Q', '173-643, 'R', '173-644, 'Q', '173-645, 'R', '173-646, 'Q', '173-647, 'R', '173-648, 'Q', '173-649, 'R', '173-650, 'Q', '173-651, 'R', '173-652, 'Q', '173-653, 'R', '173-654, 'Q', '173-655, 'R', '173-656, 'Q', '173-657, 'R', '173-658, 'Q', '173-659, 'R', '173-660, 'Q', '173-661, 'R', '173-662, 'Q', '173-663, 'R', '173-664, 'Q', '173-665, 'R', '173-666, 'Q', '173-667, 'R', '173-668, 'Q', '173-669, 'R', '173-670, 'Q', '173-671, 'R', '173-672, 'Q', '173-673, 'R', '173-674, 'Q', '173-675, 'R', '173-676, 'Q', '173-677, 'R', '173-678, 'Q', '173-679, 'R', '173-680, 'Q', '173-681, 'R', '173-682, 'Q', '173-683, 'R', '173-684, 'Q', '173-685, 'R', '173-686, 'Q', '173-687, 'R', '173-688, 'Q', '173-689, 'R', '173-690, 'Q', '173-691, 'R', '173-692, 'Q', '173-693, 'R', '173-694, 'Q', '173-695, 'R', '173-696, 'Q', '173-697, 'R', '173-698, 'Q', '173-699, 'R', '173-700, 'Q', '173-701, 'R', '173-702, 'Q', '173-703, 'R', '173-704, 'Q', '173-705, 'R', '173-706, 'Q', '173-707, 'R', '173-708, 'Q', '173-709, 'R', '173-710, 'Q', '173-711, 'R', '173-712, 'Q', '173-713, 'R', '173-714, 'Q', '173-715, 'R', '173-716, 'Q', '173-717, 'R', '173-718, 'Q', '173-719, 'R', '173-720, 'Q', '173-721, 'R', '173-722, 'Q', '173-723, 'R', '173-724, 'Q', '173-725, 'R', '173-726, 'Q', '173-727, 'R', '173-728, 'Q', '173-729, 'R', '173-730, 'Q', '173-731, 'R', '173-732, 'Q', '173-733, 'R', '173-734, 'Q', '173-735, 'R', '173-736, 'Q', '173-737, 'R', '173-738, 'Q', '173-739, 'R', '173-740, 'Q', '173-741, 'R', '173-742, 'Q', '173-743, 'R', '173-744, 'Q', '173-745, 'R', '173-746, 'Q', '173-747, 'R', '173-748, 'Q', '173-749, 'R', '173-750, 'Q', '173-751, 'R', '173-752, 'Q', '173-753, 'R', '173-754, 'Q', '173-755, 'R', '173-756, 'Q', '173-757, 'R', '173-758, 'Q', '173-759, 'R', '173-760, 'Q', '173-761, 'R', '173-762, 'Q', '173-763, 'R', '173-764, 'Q', '173-765, 'R', '173-766, 'Q', '173-767, 'R', '173-768, 'Q', '173-769, 'R', '173-770, 'Q', '173-771, 'R', '173-772, 'Q', '173-773, 'R', '173-774, 'Q', '173-775, 'R', '173-776, 'Q', '173-777, 'R', '173-778, 'Q', '173-779, 'R', '173-780, 'Q', '173-781, 'R', '173-782, 'Q', '173-783, 'R', '173-784, 'Q', '173-785, 'R', '173-786, 'Q', '173-787, 'R', '173-788, 'Q', '173-789, 'R', '173-790, 'Q', '173-791, 'R', '173-792, 'Q', '173-793, 'R', '173-794, 'Q', '173-795, 'R', '173-796, 'Q', '173-797, 'R', '173-798, 'Q', '173-799, 'R', '173-800, 'Q', '173-801, 'R', '173-802, 'Q', '173-803, 'R', '173-804, 'Q', '173-805, 'R', '173-806, 'Q', '173-807, 'R', '173-808, 'Q', '173-809, 'R', '173-810, 'Q', '173-811, 'R', '173-812, 'Q', '173-813, 'R', '173-814, 'Q', '173-815, 'R', '173-816, 'Q', '173-817, 'R', '173-818, 'Q', '173-819, 'R', '173-820, 'Q', '173-821, 'R', '173-822, 'Q', '173-823, 'R', '173-824, 'Q', '173-825, 'R', '173-826, 'Q', '173-827, 'R', '173-828, 'Q', '173-829, 'R', '173-830, 'Q', '173-831, 'R', '173-832, 'Q', '173-833, 'R', '173-834, 'Q', '173-835, 'R', '173-836, 'Q', '173-837, 'R', '173-838, 'Q', '173-839, 'R', '173-840, 'Q', '173-841, 'R', '173-842, 'Q', '173-843, 'R', '173-844, 'Q', '173-845, 'R', '173-846, 'Q', '173-847, 'R', '173-848, 'Q', '173-849, 'R', '173-850, 'Q', '173-851, 'R', '173-852, 'Q', '173-853, 'R', '173-854, 'Q', '173-855, 'R', '173-856, 'Q', '173-857, 'R', '173-858, 'Q', '173-859, 'R', '173-860, 'Q', '173-861, 'R', '173-862, 'Q', '173-863, 'R', '173-864, 'Q', '173-865, 'R', '173-866, 'Q', '173-867, 'R', '173-868, 'Q', '173-869, 'R', '173-870, 'Q', '173-871, 'R', '173-872, 'Q', '173-873, 'R', '173-874, 'Q', '173-875, 'R', '173-876, 'Q', '173-877, 'R', '173-878, 'Q', '173-879, 'R', '173-880, 'Q', '173-881, 'R', '173-882, 'Q', '173-883, 'R', '173-884, 'Q', '173-885, 'R', '173-886, 'Q', '173-887, 'R', '173-888, 'Q', '173-889, 'R', '173-890, 'Q', '173-891, 'R', '173-892, 'Q', '173-893, 'R', '173-894, 'Q', '173-895, 'R', '173-896, 'Q', '173-897, 'R', '173-898, 'Q', '173-899, 'R', '173-900, 'Q', '173-901, 'R', '173-902, 'Q', '173-903, 'R', '173-904, 'Q', '173-905, 'R', '173-906, 'Q', '173-907, 'R', '173-908, 'Q', '173-909, 'R', '173-910, 'Q', '173-911, 'R', '173-912, 'Q', '173-913, 'R', '173-914, 'Q', '

A;Contents: annotation: X-ray crystallography, 1.9 angstroms
C;Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
C;Genetics:
A;Gene: GFP
A;Introns: 69/3; 167/3
A;Superfamily: green-fluorescent protein
C;Keywords: chromoprotein; luminescence
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.0%; Score 1236; DB 1; Length 238;
Best Local Similarity 96.6%; Pred. No. 5.1e-96;
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 VSKGEELFTGVVPIVLVDGVNGHKFSVSCEGEGDATYGLTKLTKFTCTTGKLPVWPVL 61
Db 1 MSKGEELFTGVVPIVLVDGVNGHKFSVSCEGEGDATYGLTKLTKFTCTTGKLPVWPVL 60

Qy 62 VTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTLV 121
Db 61 VTTSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTLV 120

Qy 122 NRIELKGIDFKEDGNILGHKLEYNYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

Qy 182 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLAFVTAAGITLGMDELYK 239
Db 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLAFVTAAGITLGMDELYK 238

RESULT 2
H72228
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: H72228
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-785 <ARN>
A;Cross-references: GB:AE001806; GB:AE000512; NID:G4982196; PIDN:AAD36691.1; PID:G498219
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1624

Query Match 8.3%; Score 106; DB 2; Length 785;
Best Local Similarity 19.7%; Pred. No. 0.75;
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

Qy 3 SKGEELFTGVVPIVLVDGVNGHKFSVSCEGEGDATYGLTKLTKFTCTTGKLPVWPVL 62
Db 15 NEGRTSFEGTVPGVQAD-----LVKRGLLPHPYGM- 46

Qy 63 TTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTLVN 122
Db 47 -----NEDLFKEIDREWIEYEFKEDVKEGVRDLVFEVDTL 88

Qy 123 RIELKGIDFKEDGNILGHKLEYNYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVOLADH 182
Db 89 DVTLVNGVYL---GSTEDMEIYRFDVTNL-----KEKNHLKYIK-----SPIRVPKT 134

Qy 183 YQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLAFVTAAGITLGM 235
Db 135 LEQNYGLVGGP-----EDF-----IRGYRKAQSYGWD 163

RESULT 3
H64102
leucine-CRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)
N;Alternate names: leucyl-tRNA synthetase
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
A;Accession: H64102
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C.
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: H64102
A;Molecule type: DNA
A;Residues: 1-861 <TIGR>
A;Cross-references: GB:U32774; GB:L42023; NID:G1573942; PIDN:AAC22581.1; PID:G1573943; T;
C;Genetics:
A;Gene: leuS
C;Superfamily: leucine-tRNA ligase
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.2%; Score 91.5; DB 2; Length 861;
Best Local Similarity 24.1%; Pred. No. 14;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

Qy 50 TTGKLPVWPVLTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGN 103
Db 314 TGDKLPVWPVLMVNFVLMHGYTGAVMAVPAH-DQDFF-----EFAQKYSLPKQVIAPIA 364

Qy 104 DGNKYKTRAEVKEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNHNHVIIMADK-QKNGI 162
Db 365 DESIDITKQAFVHGHKLVNSDEFDGKPF--DAFNG-----IADKLEKLV 408

Qy 163 ---KNFKIRH-----NIEDGSVOLADHYQONTPTGDPVLLPDNHYL- 202
Db 409 GKQVNYRLRDMGVSQRQYWGAPIPMLTLENGDVVPA-----PMEDLPILPEDVVM 461

Qy 203 STQSALSKDPN 213
Db 462 GVKSPINADFN 472

RESULT 4
T06586
DNA-binding protein PD2 - garden pea
C;Species: Pisum sativum (garden pea)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C;Accession: T06586
R;Sato, N.; Kazuno, A.A.; Ohta, N.; Ohshima, K.
submitted to the EMBL Data Library, June 1996
A;Description: Identification of a novel family of DNA-binding proteins with two AT-hook
A;Reference number: Z15774
A;Accession: T06586
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-632 <SAT>
A;Cross-references: EMBL:X98740; NID:E995229; PIDN:CAA67292.1; PID:E275185
A;Experimental source: cv. Alaska

Query Match 7.1%; Score 90; DB 2; Length 632;
Best Local Similarity 23.3%; Pred. No. 12;
Matches 49; Conservative 26; Mismatches 79; Indels 56; Gaps 7;

Qy 16 LVLEDGVNGHKFSVSCEGEGDATYGLTKLTKFTCTTGKLPVWPVLTTLSYGVQCF 75
Db 363 IVQGRDVGSKVDVINKSNEATIPENK-----PTEPKLDVQEQLAATM----- 408

Qy 76 PDHMKQHDFFKSAMPEGYV-----QERTIFFKDDGNKYKTRAEVKEGDTLVNRIE 125

Db 409 PSSAKNVLTLDLIVFTPLRSVARTSGREGSEELKDGNSLERDTKKLELEQGKNS-E 467

QY 126 LKGDIFKEDGNILGHKLE-----NYNSHNVYIMADKOK-----N 160

Db 468 LKGIETDSTLDEKFEFENALGNKILKEISNPRHDVESANSHTHNQVTVSHQKALETNW 527

QY 161 GIKVNFKIRHNIBDG-----SVQLADHYQ 194

Db 528 QSQVEDVAKNIQDDSKPSESLHAKDYR 557

RESULT 5

E82590

leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C:Accession: E82590

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: E82590

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-887 <SIM>

A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigh chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.D.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, B.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2176

C:Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 887;

Best Local Similarity 22.2%; Pred. No. 21;

Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPTLVTLSYGVCFSRYPDHMKQHDFFKSAPEGYQVETIFFKDDGNY-- 107

Db 329 TNEQLPV-VVNFVLMAYGTGAVMAYEGHDQRQDEF--ANKYGLPIRQVIALKEPKNQDE 385

QY 108 -----KTRAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYNVSHVYI 153

Db 386 STWEPDWRDWDYAKTR---EFE---LINSAEFDGLDYQDAEVLAEFE----- 429

QY 154 MADKQKNG-IKVNFKIRHNIEDSGVQLADHYQNTPI-----GGPVLPLPN 199

Db 430 ---RQGRQRVNYRLR----DWGVSQRVWGCPFVIVYPTCGAVPVPEDQLPVILPEN 482

QY 200 -HYLSTQSALSCKDPNEKR 216

Db 483 VAFSGTGSPKTPDPEWRK 500

RESULT 6

I40794

dihydrolipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum

N;Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase hydrogense complex chain E3; S-complex 50K chain

C:Species: Clostridium magnum

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: I40794

R;Kruger, N.; Opermann, F.B.; Lorenzl, H.; Steinbuchel, A.

J;Bacteriol. 176, 3614-3630, 1994

A:Title: Biochemical and molecular characterization of the Clostridium magnum acetoin de

A:Reference number: I40789; PMID:94266715; PMID:8206840

A:Accession: I40794

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-578 <RU>

A:Cross-references: GB:IJ1844; NID:g472324; PIDN:AAA21748.1; PID:g472330

C:Function:

A:Description: catalyzes the oxidation of dihydrolipoamide to lipoamide using NAD

A:Pathway: acetoin dehydrogenase enzyme system

C:Superfamily: Alkaligenes dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase

C:Keywords: FAD; flavoprotein; lipoamide; NAD; oxidoreductase; redox-active disulfide

F15-77/Domain: lipoyl/biotin-binding homology <LFB>

F117-145/Region: beta-alpha-beta FAD nucleotide-binding fold

F119-561/Domain: dihydrolipoamide dehydrogenase homology <DLB>

F127-315/Region: beta-alpha-beta NAD nucleotide-binding fold

F153-158/Disulfide bonds: redox-active #status predicted

Query Match 7.0%; Score 89; DB 1; Length 578;

Best Local Similarity 22.9%; Pred. No. 13;

Matches 57; Conservative 43; Mismatches 91; Indels 58; Gaps 13;

QY 10 TGVPTLVLDGVDNGHKFSVSGEGDGYTKLTK-----FICTGKLPVWPPTLVT 63

Db 255 TGSMPFIPDIE-----GNKLS-----GUIDST-GALSLESNPESIALIGGVIGVEFASIFN 305

QY 64 TLSYGVCFSRYPDHMKQHDFFKSAPEGYQVETIFFKDDGNYKTRAEVKFEGDTLVNR 123

Db 306 SLGCKSVIIEMLPHILPPMDREISEI-----AKAKLIRDGININN 346

QY 124 IELKGIDFKEDG---NILGHKLEYNVSHVIMADKQK---GIKNFKIRHNIEDGSVQ 178

Db 347 CKVTRIEQEDGLKVSFIDGKEESIDVEKVLIAVGRSNIIEGLDVE-KIGVKTEGSGII 405

QY 179 LADHYQONT---PIGD--GPVLLPDNHYLSTQSALSCKD---PNEKRHDHMLAGFV--- 225

Db 406 VNDKMETNVEGIYALGDCTGKIMLA--HVASQGVVAENIMQNKMDYKTVPCVYTK 463

QY 226 ---TAAGIT 231

Db 464 PELASVGLT 472

RESULT 7

D83917

DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: D83917

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hirano

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; PMID:20512592; PMID:11058132

A:Accession: D83917

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-655 <STO>

A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05859.1; GSPDB:GN001

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2140

C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 6.9%; Score 88.5; DB 2; Length 655;

Best Local Similarity 21.9%; Pred. No. 17;

Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK---FSVSGEGDGT----YGLTKLTKFI-----CTTGKLPVWP 59

```
Db      63  NVTIHKDQSVSRDEGRGMPGTMHKLKPTPEVILTVLHAGGKFGGGVATSGGLHGVGA 122
QY      60  TLVTTLSYGVCFSRYPDHMKQHDFFKSAAMPEGYVQER-----TIFPKDGD----- 105
Db      123  SVYNALSEWLVIRKRDGWYEQRFENGKPGSTTLEKKKGTROTGTTHFKPDPVTFSTT 182
QY      106  --NFKRAEVKFGDGLVNRRIELKGDIDFKEDGNILGHKLEYNVSHNVYIMADK----- 157
Db      183  NFNVETLSELRRAAFLLKGLKLELDVLRDDTKEVPH- YEDGKA FVEYVLENEDEKTELHPV 241
QY      158  -----QKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGFVLLPNHNYLSTQSALSKDP 212
Db      242  VFENGESNGLEIEFAQFN--DGYTENVLFSVNVNRTKDG-----GTHELGAKTANTRAV 294
QY      213  NE 214
Db      295  NE 296

RESULT 8
E70390
Iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus
N;Contains: L-cysteine sulfotransferase (EC 2.8.1.-)
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999
C;Accession: E70390
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70390
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-370 <AQF>
A;Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065
A;Experimental source: strain VFS
C;Genetics:
C;Superfamily: nitrogen fixation protein nifs
C;Keywords: phosphoprotein; pyridoxal phosphate; sulfotransferase
F;195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted
F;318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match      6.9%; Score 87.5; DB 2; Length 370;
Best Local Similarity 25.4%; Pred. No. 9.9;
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY      4  KGBELFTGVV-----PILVELD----GDVNGHKE-SVSGEG----EGDATYGLKLTLPFICT 50
Db      164  KGVPLLTDAVQATGKPIELKNISYATFSKPKHAKGSLFLISDEANVEPLIVGGQE 223
QY      51  TGKLP-----VPMPTLTVTLLSYGVQCFSRYPDHMKQ-HDFFKSAMPEGYVQERTIFFPKOD 104
Db      224  NGRKSGTENNVGILSLAKALEIIVNFSRYOEQLKRLDLFENLLLEA-LPDAQIVGKDA 282
QY      105  GNVKTRAEV---KPEGDTLVNRRIELKGDIDFKEDGNILGHKLEYNVSHNVYIMADKQNG 151
Db      283  ERSFSSIVIMPFFGAEIYNKLSKGIYCTSGSACLSGSEYEPNKMFKMGFSQEXALRM 342
QY      162  IKVNFKIRHNIED 174
Db      343  VRFSFGLLNKEE 355

RESULT 9
AC0582
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0582
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
```

```
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0582
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-860 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176
C;Genetics:
A;Gene: STY0699
C;Superfamily: leucine-tRNA ligase

Query Match      6.9%; Score 87.5; DB 2; Length 860;
Best Local Similarity 23.3%; Pred. No. 30;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY      50  TTCKLPVPMPTLVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFPKDDGNYKT 109
Db      314  TGEIIPV-WAANFVLMYEYGTGAVMVPGH-DORD-YEFASKYGLTIKPVILAADGSEPD 370
QY      110  RAEVKEGDTLVNRRIELKGDIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIR 169
Db      371  SEQALTEKGVLFNSGEGFDGLAFEAFAFADKL-----AEKGVGERKVNRLR 418
QY      170  H-----NTEGDSVQLADHYQQNTPIGDGFVLLPNHNYL-STQSALSKDP 212
Db      419  DMGVSQRQYWGAPIPMVTLEDGTV-----LPTPDQLPEVLDPEDVMDGITSPKADP 471

RESULT 10
A36028
DNA-directed DNA polymerase (BC 2.7.7.7) II catalytic chain - yeast (Saccharomyces cerevisi
N;Alternate names: DNA-directed DNA polymerase II chain A; protein N0825; protein YNL262a
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: A36028; B36028; S60919; S63235; S65121
R;Morrison, A.; Araki, H.; Clark, A.B.; Hamatake, R.K.; Sugino, A.
Cell 62, 1143-1151, 1990
A;Title: A third essential DNA polymerase in Saccharomyces cerevisiae.
A;Reference number: A36028; MUID:90381771; PMID:2169349
A;Accession: A36028
A;Molecule type: DNA
A;Residues: 1-2222 <MOR>
A;Cross-references: GB:M60416; GB:M36724; NID:gl71408; PIDN:AAA88711.1; PID:gl71409
A;Accession: B36028
A;Molecule type: protein
A;Residues: 1214-1216, 'X', 1218-1221 <MO2>
R;Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.M.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV fr
A;Reference number: S60909
A;Accession: S60919
A;Molecule type: DNA
A;Residues: 1-2221 <SEN>
A;Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247
R;Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63235
A;Accession: S63235
A;Molecule type: DNA
A;Residues: 1-2222 <SEW>
A;Cross-references: EMBL:Z71538; NID:gl302316; PIDN:CAA96169.1; PID:gl302317; GSPDB:GN000
A;Experimental source: strain S288C
R;Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.
Yeast 12, 505-514, 1996
A;Title: The sequence of a 24 152 bp segment from the left arm of chromosome XIV from Sac
A;Reference number: S65111; MUID:96310631; PMID:8740425
A;Accession: S65121
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
```


A:Residues: 1-2221 <SEF>
A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C:Genetics:
A:Gene: SGD:POL2; DUN2; MIPS:YNL262w
A:Cross-references: SGD:S0005206; MIPS:YNL262w
A:Map position: 14L
C:Superfamily: DNA-directed DNA polymerase II
C:Keywords: DNA binding; nucleotidyltransferase; nucleus; zinc finger

Query Match 6.9%; Score 87.5; DB 1; Length 2222;
Best Local Similarity 28.2%; Pred. No. 1e+02;
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;

QY 54 LPVPWP-TLVTTLSYGVCFSRYPDRH-----KQHDFFKSAPEGVV-----QERTI 99
DB 883 LPKSPETVFTLENGKLLYSPCSMLNRYVHOKETNQYQELKDPLNYIVYTHSENTI 942
QY 100 FFKDDGNYKTR--AEVKFGDGLVNR-----IELKGIDFKBDGNILGHKLENNY 147
DB 943 FFEVDGPYKAMILPSSKEGKIKRYAVFNEGDSLAELKGFELKRRGEL---QLIKNFQ 999
QY 148 S--HNYYIMAD 156
DB 1000 SDIFKVFLGED 1010

RESULT 11
D71614
hypothetical protein PFB0460c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
R:Accession: D71614
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Perrea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: D71614
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2573 <GAR>
A:Cross-references: GB:AE001396; GB:AE001362; NID:g3845188; PIDN:AAC71881.1; PID:g384519
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;
Best Local Similarity 26.2%; Pred. No. 1.2e+02;
Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VQERTIFFKD--DGNVYKTRAEVKFEGDGLVNRLELKGIDFKEDGNILGHKLENN--YNH 149
DB 126 LKKEITLKDIKSGSDNPDEISLFDKDDWDDKELK--DFEKSLLKKNKEVNFYNNKM 183
QY 150 NVYIMADKQNGIKVNFKRHNIEDGSGVOLADHYQONTPIGDGPVLLPDNHYLSQSALS 209
DB 184 NLHIKENKKDEKKNKHNDENNMM-----IYYKNI---DKTHYLDNNVVHILNDIN 236
QY 210 KDPNEKRDHM 219
DB 237 TYLKRERDYM 246

RESULT 12
G81355
tRNA (uracil-5-)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter jejuni
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
R:Accession: G81355
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: G81355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <PAR>
A:Cross-references: GB:ALJ39076; GB:ALJ11168; NID:g6968128; PIDN:CAB73096.1; PID:g696827;
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: trmA; Cj0831c
C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 6.8%; Score 86; DB 2; Length 357;
Best Local Similarity 24.8%; Pred. No. 13;
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;

QY 80 KOHDFKSAPEGVVQERTIFFKDDGNYKTRAEVKF--EGDGLV-----NRIELKG 138
DB 14 EKHSFIKFKFEFYTKDFKLFASKDKHYTRAELSFYHENDTLFYAMFDPKSKKKYIIIEY 73
QY 129 IDFKED-----GNILGHKLENNYNSHNYYIMADKQNGIKVNFKRHNIE 173
DB 74 LDFADEKICAFMRLLLEYLQRDNKLKEKL-----FGVEFLTKQE--LSITLLYHKNIE 125
QY 174 D 174
DB 126 D 126

RESULT 13
T06936
photosystem II chlorophyll a-binding protein psbC - Cyanophora paradoxa cyanelle
N:Alternate names: protein CP43
C:Species: cyanelle Cyanophora paradoxa
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 26-Aug-1999
R:Accession: T06936
R:Stirewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A:Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A:Reference number: Z15840
A:Accession: T06936
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-461 <STI>
A:Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81279.1; PID:g1016192
A:Experimental source: strain Pringsheim LB555
C:Genetics:
A:Gene: psbC
A:Genome: cyanelle
C:Superfamily: photosystem II chlorophyll a-binding protein psbC
C:Keywords: chlorophyll; cyanelle; membrane-associated complex; photosynthesis; photosyst

Query Match 6.8%; Score 86; DB 2; Length 461;
Best Local Similarity 21.0%; Pred. No. 18;
Matches 57; Conservative 35; Mismatches 89; Indels 90; Gaps 13;

QY 25 GHKFSVSGEGEDATYKGLTLKFICTGKL-----PVPWP--TLVTT-----LSYGV--- 69
DB 208 GGGWIVSDNMEDIGGHIWLAFCIIGVWHILTKFPFWARRALVWSGEAYLSYSIAL 267
QY 70 -----QCFSRYPDHMKOHDFKFSAMPE-GYVQERTIFFKOD-----GN 106
DB 268 ALMGFIANCFVNFNTAYPSEFFGPTGPASQAQFTFLVDRQLGANVGSQAQPTGLGK 327
QY 107 YKTRA---BVKFEGDGLVNRLELKGIDFKEDGNILGHKLENNYNSHNYYIMADKQNGIK 163
DB 328 YLMRSPSGEIIIFGGETM-----RFDTRAPWLEPLRGANGLD 364
QY 164 VNFKRHNIEDGSGVOLADHYQONTPIGD---GPVLLPDN--HYLSQSALS KDPNEKRD 217
DB 365 LT-KIKYIQPWOERRAABYMTAPUGSLNSVGGVATEINSVYSPRSWLS-----TS 417
QY 218 HMVLAGFV-----TAAGITLGMD 235

```

487 FFENYDMSKSDTSSNYKRTTYGNSVTL-GFPVNNNSYYVGLGHTYKNISNFALYN---542
QY HNVYIMADKQK-NGIKVNFKIRINIEDGSQLADHYQQ-----NTPIGDGPVLL196
DB RLNYIOSMKFKNGIKTN-----DFDFSGWNTNSLNRGFFPTKGVKASLG-GRVTI593
QY P--DNHYLSQTQSALSKDPNEKRDHMLVLAGFVTTAAGITLG233
DB PGSDNKYIKLSADVQGFYPLDRDHLWVWSAKASAGYANG632

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-222a

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVPILVELD.....VLAFVTAAGITLGMDELK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1242	97.5	238	1	GFP_AEQVI
2	91.5	7.2	861	1	SYL_HAEN
3	89.5	7.0	879	1	SYL_XYLFA
4	87.5	6.9	860	1	SYL_SALTI
5	87.5	6.9	860	1	SYL_SALTY
6	87.5	6.9	2222	1	DPOE_YEAST
7	87	6.8	689	1	AC2L_HUMAN
8	86.5	6.8	533	1	CP51_CANGA
9	86.5	6.8	879	1	SYL_XYLFT
10	86.5	6.8	1603	1	VIT4_CAEEL
11	86	6.8	357	1	TRMB_CAMJE
12	86	6.8	461	1	PSBC_CYAPA
13	86	6.8	501	1	YC03_KLEPN
14	85.5	6.7	501	1	AMPA_WICBR
15	85.5	6.7	788	1	DPOL_HPBHE
16	85.5	6.7	795	1	D152_HAEN
17	85.5	6.7	737	1	D151_HAEN
18	85.5	6.7	886	1	ITH3_MESAU
19	84.5	6.6	589	1	SYD_HAEDU
20	84.5	6.6	613	1	PEPF_MYCPU
21	84.5	6.6	793	1	D153_HAEN
22	84.5	6.6	859	1	SYL_SHRON
23	84.5	6.6	941	1	GUN_BACS6
24	84	6.6	353	1	HIS7_BUCAI
25	84	6.6	366	1	SET7_HUMAN
26	84	6.6	874	1	SLAP_BACLI
27	83.5	6.6	533	1	NIFD_CLOPA
28	83.5	6.6	538	1	GRBE_RAT
29	83.5	6.6	1164	1	BAG_STRAG
30	82	6.4	336	1	YD4B_METJU
31	82	6.4	461	1	PSBC_CHLEU
32	82	6.4	682	1	PRC_ECOLI
33	82	6.4	737	1	OPTI_DROME

RESULT 1
GFP_AEQVI STANDARD; PRT; 238 AA.
AC P42212; Q17104; Q27903;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92175527; PubMed=1347277;
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA Cormier M.J.;
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein.";
RL Gene 111:229-233(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94185810; PubMed=8137953;
RA Inouye S., Tsuji F.I.;
RT "Aequorea green fluorescent protein. Expression of the gene and
RT fluorescence characteristics of the recombinant protein.";
RL FEBS Lett. 341:277-280(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RX Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RA "Enhanced expression in tobacco of the gene encoding green fluorescent
RT protein by modification of its codon usage.";
RL Plant Mol. Biol. 33:989-999(1997).
RN [4]
RP CHROMOPHORE.
RX MEDLINE=93192221; PubMed=8448132;
RX Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;
RA "Chemical structure of the hexapeptide chromophore of the Aequorea
RT green-fluorescent protein.";
RL Biochemistry 32:1212-1218(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=96355665; PubMed=8703075;
RX Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,
RA Remington S.J.;
RT "Crystal structure of the Aequorea victoria green fluorescent
RT protein.";
RL Science 273:1392-1395(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=98294543; PubMed=9631087;
RA Yang F., Moss L.G., Phillips G.N. Jr.;
RT "The molecular structure of green fluorescent protein.";
RL Nat. Biotechnol. 14:1246-1251(1996).

ALIGNMENTS

34	82	6.4	752	1	NEC1_RAT	P28840	rattus norv
35	81.5	6.4	393	1	TRMB_HELPY	O25443	helicobacte
36	81	6.4	682	1	AC2L_MOUSE	Q99ub1	mus musculu
37	81	6.4	774	1	AMY2_SCHPO	Q42918	schizosacch
38	80.5	6.3	658	1	ADAS_HUMAN	O00116	homo sapien
39	80.5	6.3	860	1	SYL_ECO57	Q8xbn8	escherichia
40	80.5	6.3	860	1	SYL_ECOL6	Q8fjy9	escherichia
41	80.5	6.3	860	1	SYL_ECOLI	P07813	escherichia
42	80.5	6.3	1259	1	YTFN_ECOLI	P39321	escherichia
43	80	6.3	439	1	SY62_DISOM	P24506	discofyge o
44	80	6.3	501	1	MTBB_BACSU	P33563	bacillus eu
45	80	6.3	728	1	CATB_ASPPU	Q92405	aspergillus

[7]
 X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
 MEDLINE-98455509; PubMed-9782051;
 Wachter R.M., Eislinger M.A., Kallio K., Hanson G.T., Remington S.J.;
 "Structural basis of spectral shifts in the yellow-emission variants
 of green fluorescent protein.";
 Structure 6:1267-1277(1998).
 [8]
 X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 MEDLINE-99238303; PubMed-10220315;
 Eislinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
 "Structural and spectral response of green fluorescent protein
 variants to changes in pH";
 Biochemistry 38:5296-5301(1999).
 -1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the
 blue chemiluminescence of the protein aequorin into green
 fluorescent light by energy transfer. Fluoresces in vivo upon
 receiving energy from the Ca(2+)-activated photoprotein aequorin.
 Absorbs light maximally at 395 nm and exhibits a smaller
 absorbance peak at 470 nm. The fluorescence emission spectrum
 peaks at 509 nm with a shoulder at 540 nm.
 -1- SUBUNIT: Monomer.
 -1- TISSUE SPECIFICITY: Photocytes.
 -1- PTM: Contains a covalently attached chromophore, which is composed
 of modified amino acid residues. The chromophore is formed upon
 cyclization of the residues Ser-dehydroTyr-Gly.
 -1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
 chimeric proteins of GFP linked to other proteins where it
 functions as a fluorescent protein tag. GFP tolerates N- and C-
 terminal fusion to a broad variety of proteins. It has been
 expressed in bacteria, yeast, slime mold, plants, Drosophila,
 zebrafish, and in mammalian cells. As a noninvasive fluorescent
 marker in living cells, it allows for a wide range of applications
 where it may function as a cell lineage tracer, reporter of gene
 expression, or as a measure of protein-protein interactions.
 -1- DATABASE: NMR-Database; Protein Spotlight;
 NOTE=Issue 11 of June 2001;
 WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; MG2654; AAA27722.1; -;
 EMBL; MG2653; AAA27721.1; -;
 EMBL; L29345; AAA58246.1; -;
 EMBL; X96418; CAA65278.1; -;
 FIC; J30692; JQ1514.
 DR PDB; 139C; 17-NOV-00.
 DR PDB; 13FP; 07-JUL-97.
 DR PDB; 1C4F; 14-JUN-00.
 DR PDB; 1EMA; 08-NOV-96.
 DR PDB; 1EMB; 16-JUN-97.
 DR PDB; 1EMC; 20-AUG-97.
 DR PDB; 1EME; 20-AUG-97.
 DR PDB; 1EMF; 20-AUG-97.
 DR PDB; 1EMG; 12-MAY-99.
 DR PDB; 1EMK; 20-AUG-97.
 DR PDB; 1EMI; 20-AUG-97.
 DR PDB; 1EMW; 20-AUG-97.
 DR PDB; 1F03; 17-NOV-00.
 DR PDB; 1F0B; 17-NOV-00.
 DR PDB; 1GFL; 11-JAN-97.
 DR PDB; 1HCU; 15-JAN-02.
 DR PDB; 1HUY; 04-JUL-01.
 DR PDB; 1JBY; 07-JAN-03.
 DR PDB; 1JBZ; 07-JAN-03.
 DR PDB; 1KPS; 28-AUG-02.
 DR PDB; 1KYP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.
 DR PDB; 1KYS; 10-APR-02.
 DR PDB; 1YFP; 28-OCT-98.
 DR PDB; 2EMD; 20-AUG-97.
 DR PDB; 2EMN; 20-AUG-97.
 DR PDB; 2EMO; 20-AUG-97.
 DR InterPro: IPR009017; GFP like.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS; PR01229; GFLUORESCENT.
 DR ProDom; PD013756; Green fl protein; 1.
 KW Luminescence; 3D-structure.
 FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).
 FT MOD_RES 66 66 2,3-DIDEHYDROTYROSINE.
 FT VARIANT 100 100 F -> Y.
 FT VARIANT 108 108 T -> S.
 FT VARIANT 141 141 L -> M.
 FT VARIANT 219 219 V -> I.
 FT CONFLICT 2 2 S -> G (IN REF. 3).
 FT CONFLICT 25 25 H -> Q (IN REF. 2).
 FT CONFLICT 80 80 Q -> R (IN REF. 3).
 FT CONFLICT 157 157 Q -> P (IN REF. 2).
 FT CONFLICT 172 172 E -> K (IN REF. 2).
 FT HELIX 4 8
 FT STRAND 12 22
 FT TURN 23 24
 FT STRAND 25 36
 FT TURN 37 40
 FT STRAND 41 48
 FT TURN 49 50
 FT HELIX 57 60
 FT TURN 61 63
 FT HELIX 69 71
 FT STRAND 73 73
 FT HELIX 76 81
 FT HELIX 83 86
 FT TURN 87 90
 FT STRAND 92 100
 FT TURN 101 102
 FT STRAND 105 115
 FT TURN 116 117
 FT STRAND 118 128
 FT TURN 132 133
 FT TURN 135 139
 FT STRAND 141 141
 FT STRAND 148 155
 FT TURN 156 159
 FT STRAND 160 171
 FT TURN 172 173
 FT STRAND 176 187
 FT STRAND 199 208
 FT TURN 211 212
 FT STRAND 217 227
 SQ SEQUENCE 238 AA; 2686 MW; EASA6F21PFB6E05 CRC64;
 Query Match 97.5%; Score 1242; DB 1; Length 238;
 Best Local Similarity 97.9%; Pred No. 5.5e-95;
 Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 VSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGLTLKPICTTGKLPVWPPTL 61
 :|||||
 Db 1 MSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGLTLKPICTTGKLPVWPPTL 60
 QY 62 VTTLVSGVQCFSRYPDHMKQHDFFKSAPEGVVOERTIFFKDDGNKTRAEVKFEGDTLV 121
 :|||||
 Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAPEGVVOERTIFFKDDGNKTRAEVKFEGDTLV 120
 QY 122 NRIELKGIDFKEDGNILGHKLEYNVNSHVYIMADKQNGIKVNFKIRHNIEDSGVOLAD 181
 :|||||
 Db 121 NRIELKGIDFKEDGNILGHKLEYNVNSHVYIMADKQNGIKVNFKIRHNIEDSGVOLAD 180
 QY 182 HYQONTPIGDGPVILLPDNHYLSTQSALSADPKNEKDHVMVLAVFTVAAGITLGMDELYK 239
 :|||||

```
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238
RESULT 2
SYL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32774; AAC22581.1; -
CC PIR; H64102; H64102.
CC TIGR; H10921; -
CC HAMAP; MF_00049; -; 1.
DR InterPro; IPR002302; Leu-trNASynt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_f1eRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; P000985; TRNASYNTHLEU.
DR TIGRPFAM; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 7;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVMPPTLVTTLTSLGVCFSRYPDHMKQHDFFKSAMPEGYQVETIFPKD-----103
Db 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKYSLPFKQVIAPLA 364
104 DGNKYTRAEVKEFGTDLVNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADK-QKNGI 162
365 DEEIDLTKQAFVEHGLVNSDFDGNF--DGAENG-----IADKLEKLGV 408
163 ---KUNFKIRH-----NIEDSVQLADHYQQNTPIGDPVLLPDNHYL- 202
409 GKQVNYRLRDWGSQRQRYGAPIMPLTLENGDVVPA-----PMEDLPFIILPEDVYMD 461
203 STQSALSADPN 213
462 GYKSPINADPN 472
RESULT 3
SYL_XYLFA
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9P688;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.B., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho A.P., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinanci L.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Souza A.P., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF004031; AAF84975.1; ALT_INIT.
DR HAMAP; MF_00049; -; 1.
```



```
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008725; AAL19599.1; -.
CC STyGene; SG????; Leus.
CC HAMAP; MF_00049; -.
CC InterPro; IPR002302; Leu-tRNA-synt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_fIERS_edit.
CC Pfam; PF001133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRfam; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA tRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 15;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVMPPTLVTTLSYGVCFSRYPDHMKQDFFKSAPEGYVQERTIFFKDDGNYKT 109
DB 314 TGEIIPV-WAANFVLMYEGTGAVMAVFGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKEGDTLVNRILKLGDKEDNGILGHKLEYNHNVIMADKQNGIKVNFKR 169
DB 371 SEQALTEKGVLFNSGFBGDLGAFAAFNAIADKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NTEDGSVQLADHYQQNTPIGDGVLPLPDNHYL-STOSALSKDP 212
DB 419 DWGVSQRQYWGAPIPWVLTEDGTV-----LFTPEDQLVILPDVMDGITSPIKADP 471

RESULT 6
ID_DPOE YEAST STANDARD; PRT; 2222 AA.
AC P21951;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
DE polymerase II subunit A)
GN POL2 OR DUN2 OR YNL262W OR N0825.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.
RX MEDLINE=90381771; PubMed=2169349;
RA Morrison A., Araki H., Clark A.B., Hamatake R.K., Sugino A.;
RT "A third essential DNA polymerase in S. cerevisiae.";
RL Cell 62:1143-1151(1990).
RN [2]
RP SEQUENCE OF 1-2221 FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;
```

```
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT genes."
RL Yeast 12:505-514(1996).
RN [3]
RP TEMPERATURE SENSITIVE MUTANTS.
RX MEDLINE=92164663; PubMed=1537345;
RA Araki H., Ropp P.A., Johnson A.L., Johnston L.H., Morrison A.,
RA Sugino A.;
RT "DNA polymerase II, the probable homolog of mammalian DNA polymerase
RT epsilon, replicates chromosomal DNA in the yeast Saccharomyces
RT cerevisiae."
RL EMBO J. 11:733-740(1992).
CC -1- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
CC REPLICATION.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -1- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30
CC kDa, AND 29 kDa).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNITS B AND C.
CC -1- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
CC alpha, beta, gamma, delta, and epsilon which are responsible for
CC different reactions of DNA synthesis.
CC -1- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; MG04116; AAA88711.1; -.
CC EMBL; X92494; CAA63235.1; -.
CC EMBL; Z71538; CAA96169.1; -.
CC PIR; A36028; A36028.
CC GERMOnline; 143268; -.
CC SGD; S0005206; POL2.
CC GO; GO:000731; P:DNA repair synthesis; IMP.
CC InterPro; IPR006172; DNA pol B.
CC InterPro; IPR006134; DNA pol B dom.
CC InterPro; IPR006133; DNA pol B_exo.
CC Pfam; PF00136; DNA pol B; 1.
CC Pfam; PF03104; DNA pol B_exo; 1.
CC SMART; SM00486; POLBG; 1.
CC PROSITE; PS00116; DNA POLYMERASE B; FALSE NEG.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Zinc-finger; Nuclear protein.
FT ZN_FING 2108 2181 POTENTIAL.
FT VARIANT 644 644 M -> I (IN POL2-9 TS MUTANT).
FT VARIANT 710 710 P -> S (IN POL2-18 TS MUTANT).
SQ SEQUENCE 2222 AA; 255669 MW; CBCDDE2AB147D65B CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 2222;
Best Local Similarity 28.2%; Pred. No. 46;
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;

QY 54 LPVPWP-TLVTTLSYGVCFSRYPDH-----KQDFFKSAPEGYV-----QERTI 99
DB 883 LPKSPFTFTFFLENGKLYLSPCSMLNRYVHQKFTNHQYQELKDPLNYIYTHSENTI 942
QY 100 FFKDDGNYKTR--AEVKEFGDTLVNR-----IELKGIDFKEDGNILGHKLEYNV 147
DB 943 FFEVDGPYKAMILPSSKEBGKIKKRYAVFNEGDSLAEKGFELKRGEL---QLIKNFQ 999
QY 148 S--HNVYIMAD 156
DB 1000 SDIFKVFLEGD 1010
```

```

RESULT 7
AC2L_HUMAN
ID AC2L_HUMAN STANDARD; PRT; 689 AA.
AC Q9NUB1; Q8IY99; Q8N234; Q96J11; Q96UX6; Q9NU28;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Acetyl-coenzyme A synthetase 2-like, mitochondrial precursor
DE (EC 6.2.1.1) (Acetate-CoA ligase 2) (Acetyl-CoA synthetase 2).
GN ACAS2L OR KIAA1846.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehaeslaio M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.B., McConnachie L.J., McIlroy K., McMurray A.D.,
RA Milne S.A., Mistry D., Moore M.F., Mullikin J.C., Nickerson T.,
RA Oliver A., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.W.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [2]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX TISSUE=Brain, and Testis;
RC MEDLINE=23288257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins B.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg K., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.C., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
SEQUENCE OF 169-689 FROM N.A. (ISOFORM 1).
RX TISSUE=Placenta, and Tongue;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,

```

```

RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamiyara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Irie R., Sato H.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nekamura Y., Sekine M.,
RA Kikuchi H., Murakawa K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Ota T., Hayashi K., Sugiyama T., Otsuki T., Ishibashi T.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S.,
RA Kawai Y., Wakamatsu A., Kanehori K., Suzuki Y., Sugano S.,
RA Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 336-689 FROM N.A. (ISOFORM 1).
RX TISSUE=Brain;
RC MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
CC -!- FUNCTION: Converts acetate to acetyl-CoA so that it can be used
for oxidation through the tricarboxylic cycle to produce ATP and
CO(2) (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
acetyl-CoA.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
Name=2;
IsoId=Q9NUB1-1; Sequence=Displayed;
IsoId=Q9NUB1-2; Sequence=VSP_007249;
Notes=No experimental confirmation available;
-!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
-!- CAUTION: Ref.1 (CAB81884) sequence differs from that shown due to a
erroneous gene model prediction.
-!- CAUTION: Ref.3 sequence differs from that shown due to a
frameshift in position 250 and numerous sequencing errors.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AL035661; CAB75500.1; -
EMBL; AL080312; CAB81884.1; ALT_SEQ.
EMBL; BC033261; AAH39261.1; -
EMBL; BC044588; AAH44588.1; -
EMBL; AK027817; BAB55390.1; ALT_INIT.
EMBL; AK092295; BAC03853.1; ALT_SEQ.
EMBL; AB058749; BAB47475.1; -
Genew; HGNC:16091; ACAS2L.
InterPro; IPR000873; AMP-bind.
Pfam; PF00501; AMP-binding; 1.
PROSITE; P500455; AMP BINDING; 1.
Ligase; Mitochondrion; Transit peptide; Alternative splicing.
TRANSIT 1 36 MITOCHONDRION (POTENTIAL).
CHAIN 37 689 ACETYL-COENZYME A SYNTHETASE 2-LIKE.
DOMAIN 45 53 POLY-ALA.
VARSP 446 447 Missing (in isoform 2).
CONFLICT 277 277 V -> M (IN REF. 2; AAH39261).
CONFLICT 488 488 V -> M (IN REF. 2; AAH44588).
SEQUENCE 689 AA; 74856 MW; 66B84E39302AD08B CRC64;
Query Match 6.8%; Score 87; DB 1; Length 689;
Best Local Similarity 24.1%; Pred. No. 13;
Matches 33; Conservative 16; Mismatches 36; Indels 52; Gaps 7;

```



```
QY 9 FTGWPFILVELDGDVNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVWPMTLVTTLSYG 68
Db 473 FFGIVPVLMDKGSV-----VEGNSVSGALCIS-----QAWPGMARTI--- 510
QY 69 VQCFSPYDPHMKQDFFKAMPBGYVQERTIPFKDDGNKYTRA---EVKFGDGLVNRIE 125
Db 511 -----YGDHQRFVDYAFKAYP-GY-----YFTGDGAYRTEGGYYQITGRMDVDVI----- 553
QY 126 LKGIDFKEDGNILGHKL 142
Db 554 -----NISGHRL 560

RESULT 8
CP51_CANGA STANDARD; PRT; 533 AA.
ID CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYP51) (P450-LIA1) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (Lia1) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -I- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -I- CATALYTIC ACTIVITY: Obtusifolliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -I- PATHWAY: Ergosterol biosynthesis.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L40389; AAB02329.1; -
DR EMBL; S75389; AAB32679.1; -
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;

KW Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.4%; Pred. No. 10;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;

QY 25 GHKFSVS---GEGEGDATYKGLTKFKICTTGKLPVWPMTLVTTLSYGVQCFSPYDPH--M 79
Db 109 GHEFIFNAKLADVSAEAYSHL-----TTPVFGKGVIVDCPNHRLM 149
QY 80 KQHDFFKSAM-PEGYV-----QERTIPFKDDGNKYTRAEVKFGDGLTVNRIELKGIDF 131
Db 150 EQKKFVKGALTKEAFVRYVPLIAEIIKYFRSKFKINENNSGIVDVNVSQPEM--TIF 207
QY 132 KEDGNILGHKLEYNYNHNVYIMADQKNGIKVKPKIRINIEDSGVOLADHQQNTPIGD 191
Db 208 TASRSLGKEMRDKLDTPAYLYSLDKGFTPINF-VFPNLPLEHYRKRDHAQAIS--- 263
QY 192 GPVLIPDNHYLSTQSALSQDPNEKD 217
Db 264 -----GTYNMSLIKERREKND 278

RESULT 9
SYL_XYLFT STANDARD; PRT; 879 AA.
ID SYL_XYLFT STANDARD; PRT; 879 AA.
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR PD1230.
OS Xylella fastidiosa (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T. da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carter H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuranae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Penille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa.";
RL J. Bacteriol. 185:1018-1026(2003).
CC -I- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```


RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 CC -!- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
 CC 54 (W-5-U54) in all tRNA (By similarity).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
 CC homocysteine + tRNA containing thymine.
 CC -!- SIMILARITY: Belongs to the RNA M5U methyltransferase family. TrmA
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL139076; CAB73096.1; -;
 DR PIR; G81355; G81355.
 DR HAMAP; MF_01011; -; 1.
 DR InterPro; IPR000051; SAM_bind.
 DR InterPro; IPR001566; TrmA.
 DR PROSITE; PS01230; TRMA 1; 1.
 DR PROSITE; PS01231; TRMA 2; FALSE NEG.
 KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
 FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY
 FT ACT_SITE 315 315 BY SIMILARITY.
 FT SEQUENCE 357 AA; 42276 MW; CEC5328347CEB497 CRC64;
 Query Match 6.8%; Score 86; DB 1; Length 357;
 Best Local Similarity 24.8%; Pred. No. 7;
 Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;
 QY 80 KOHDFKSAPEGYOERTIFFKDGNYKTRAEVKE--EGDTLV-----NRIELKG 128
 DB 14 EKHSEFKYFKFYDFDKFLASKDGHYTRAEISFYHENDILFYAMPDPKSKKVIIEY 73
 QY 129 IDPKED-----GNILGHKLEYNNSHNVIMADKQKNGIKVNFKIRHNITE 173
 DB 74 LQFADEKICAFMPRLLEYLRQDNKLEKL-----PGVEFLTKQE--LSITLLVHKITE 125
 QY 174 D 174
 DB 126 D 126
 RESULT 12
 PSBC_CYAPA STANDARD; PRT; 461 AA.
 AC P48104;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).
 GN PSBC.
 OS Cyanophora paradoxa.
 OG Cyanelle.
 OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
 OX NCBI_TaxID=2762;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UTEX LB 555 / Pringsheim;
 RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
 RA Bryant D.A.;
 RT "Nucleotide sequence of the cyanelle DNA from *Cyanophora paradoxa*.";
 RL Plant Mol. Biol. Rep. 13:327-332(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UTEX LB 555 / Pringsheim;

RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
 RA Farley J.Y., Schlachter W.M., Chung S., Newmann-Spallart C.,
 RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
 RT "The complete sequence of the cyanelle genome of *Cyanophora paradoxa*:
 RT the genetic complexity of a primitive plastid.";
 RL (in) Schenk H.E.A., Hertmann R., Jeon K.W., Mueller N.E.,
 RL Schwemmler W. (eds.);
 RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
 RL (1997).
 CC -!- FUNCTION: The 43 kDa protein (p6) is a component of the core of
 CC photosystem II. It is a chlorophyll binding protein.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Cyanelle
 CC thylakoid membrane.
 CC -!- SIMILARITY: Belongs to the psbB / psbC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U30821; AAA81279.1; -;
 DR PIR; T06936; T06936.
 DR InterPro; IPR005869; Photo44.
 DR InterPro; IPR000932; PSIIProt.
 DR Pfam; PF00421; PSII; 1.
 DR TIGRfam; TIGR01153; psbC; 1.
 KW Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Cyanelle;
 KW Transmembrane.
 SQ SEQUENCE 461 AA; 50402 MW; 62664E239E13B9C1 CRC64;
 Query Match 6.8%; Score 86; DB 1; Length 461;
 Best Local Similarity 21.0%; Pred. No. 9.4;
 Matches 57; Conservative 35; Mismatches 89; Indels 90; Gaps 13;
 QY 25 GHKFSVSGEGEDATVGLTKLFICTGKL-----PVPWP--TLVTT---LSYGV--- 69
 DB 208 GGNWISVDNMEDITGGHIWLAFLCIIGVWHLTKPFWARRALVWSEAYLSYLAAL 267
 QY 70 -----QCFSYPDHMKQHDFFKSAMPE-GYVQERTIFFKDD-----GN 106
 DB 268 ALMGFIANCFVNFNTAYPSEFFGTPPEASQAQAFTLVLRQRLCANVGSAGQPTGLGK 327
 QY 107 YKTRA---EVKPEGDTLVNRIELKGIDPKEDGNILGHKLEYNNSHNVIMADKQKNGIK 163
 DB 328 YLMRSPSGEIIIFGGETM-----RFDTRAPWLEPLRGANGLD 364
 QY 164 VNFKIRHNIEDGSLVADHYQONTPIGD---GPVLLPDN---HYLSTQSALSKDPPNEKRD 217
 DB 365 LT-KIYDIQPWQERRAAEYMTAPLGLSLNSVGGVATEINSVNVSPRSWLS-----TS 417
 QY 218 HNVLAGFV-----TAAGITLGM 235
 DB 418 HFVLGFELFIGHLHAGRAAASGGFEKGLD 448
 RESULT 13
 YC03_KLEPN STANDARD; PRT; 504 AA.
 ID YC03_KLEPN
 AC Q48479;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 55.8 kDa protein in cps region (ORF3).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Chedid;

```

RX MEDLINE=95204345; PubMed=7896702;
RA Arakawa Y., Wacharotayankun R., Nagatsuka T., Ito H., Kato N.,
RT "Genomic organization of the Klebsiella pneumoniae cps region
RT responsible for serotype K2 capsular polysaccharide synthesis in the
RT virulent strain Shedi."
RL J. Bacteriol. 177:1788-1796(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D21242; BAA04774.1; -.
KW Hypothetical protein.
SQ SEQUENCE 504 AA; 55782 MW; AD887595CFDFDFA8 CRC64;

Query Match      6.8%; Score 86; DB 1; Length 504;
Best Local Similarity 24.0%; Pred. No. 10;
Matches 46; Conservative 27; Mismatches 71; Indels 48; Gaps 9;

QY 70 QCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNKYKTRAEVKPEGD-TLVNRTELKG 128
DB 293 QSFSSFWDFGTGHD-----NCTDNEPGNQLAGDFKLEPTLGMVPSFYG 339

QY 129 IDPKED-----GNILGHKLEYNNSHNVYIMA-DKQNGIKVNFKIRHNI-BD 174
DB 340 QMWGEDESGYLPANFLGIEGHH-GWGKDAVNWYVEAHDRTNMSRTNYSYTHIYKD 398

QY 175 GSVQLADHYQONTPIGG-----PVLPPNHVLSQSALSK-DPNEKRDMHVL 221
DB 399 G-----YYQQYPLGDGMGGDQLFAGKVELITENNQRWSTRLAYAKVNPQDSINKAP 452

QY 222 AGFVTAAGITLG 233
DB 453 PHSDTLKGVLG 464

RESULT 14
ID AMPA WIGBR STANDARD; PRT; 501 AA.
AC QBD235;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase)
DE (LAP) (Leucyl aminopeptidase).
GN PEPA OR WIGBR4590.
OS Wigglesworthia glosinidia brevipalpis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Wigglesworthia.
OX NCBI_TaxID=36870;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22297718; PubMed=12219091;
RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Aksoy S.;
RT flies, Wigglesworthia glossinidia."
RL Nat. Genet. 32:402-407(2002).
CC -!- FUNCTION: Presumably involved in the processing and regular
CC turnover of intracellular proteins. Catalyzes the removal of
CC unsubstituted N-terminal amino acids from various peptides (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids
CC including Pro although not Arg or Lys, and Xbb may be Pro.
CC -!- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M17.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB063522; BAC24605.1; -.
DR HAMAP; MF_00181; -.
DR InterPro; IPR000819; Peptidase M17 C.
DR InterPro; IPR0008283; Peptidase M17_N.
DR Pfam; PF000883; Peptidase M17; 1.
DR Pfam; PF02789; Peptidase M17_N; 1.
DR PRINTS; PR00481; LAMNOPPTDASE.
DR PROSITE; PS00631; CYTOSOL_AP; 1.
KW Hydrolase; Aminopeptidase; Manganese; Complete proteome.
FT ACT_SITE 282 282 POTENTIAL.
FT ACT_SITE 356 356 POTENTIAL.
FT METAL 270 270 MANGANESE 2 (BY SIMILARITY).
FT METAL 275 275 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 293 293 MANGANESE 2 (BY SIMILARITY).
FT METAL 352 352 MANGANESE 1 (BY SIMILARITY).
FT METAL 354 354 MANGANESE 1 AND 2 (BY SIMILARITY).
SQ SEQUENCE 501 AA; 56643 MW; 4E181BBEB481PE3 CRC64;

Query Match      6.7%; Score 85.5; DB 1; Length 501;
Best Local Similarity 21.5%; Pred. No. 11;
Matches 35; Conservative 24; Mismatches 81; Indels 23; Gaps 5;

QY 93 YVOERTIPKDDGNKYKTRAEVKPEGDTLVNRTELKGIDF-KEDGNILGHKLEYNNSHNV 151
DB 153 YKKEIMFYIHDENEIK-QANIAISHSVIS-----KGIIITKLGWNPSCFDPHYLSHOS 207

QY 152 YIMADKQNGIKVNFKIRHNIEDGSVQLADH-----YQONTPIGDPVLL 196
DB 208 YILKDKYSEKISVEIMDHKKIKNGIAYLVHVSCKSNPYLSIIKYNENKFNKSPIL 267

QY 197 PDNHVLSQSALSKDPNEKRDMHV--LAGFVTAAGITLGMDEL 237
DB 268 ICKGLTFDSGGISIKPSNNMDEMCKFDMCGAAAVLGMVHAISEL 310

RESULT 15
DPOL HPBHE STANDARD; PRT; 788 AA.
ID DPOL HPBHE
AC PL3846;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P protein [Includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-
DE directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].
GN P.
OS Heron hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=28300;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88333160; PubMed=3418788;
RA Sprengel R., Kaleta E.F., Will H.;
RT "Isolation and characterization of a hepatitis B virus endemic in
RT herons."
RL J. Virol. 62:3832-3839(1988).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds
(without alignments)
2458.984 Million cell updates/sec

Title: US-09-887-784-222a

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVPIVLVD.....VLagFVTAAGITGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1238	97.2	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1235	96.9	238	5 Q93125	Q93125 aequorea vi
3	1233	96.8	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1232	96.7	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1200	94.2	238	5 Q17105	Q17105 aequorea vi
6	1185	93.0	238	5 Q17106	Q17106 aequorea vi
7	1080	84.8	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1076	84.5	238	5 Q8WP95	Q8wp95 aequorea ma
9	1072	84.1	238	5 Q8WPC4	Q8wtc4 aequorea ma
10	1070	84.0	238	5 Q8WTD0	Q8wtcd aequorea ma
11	1069	83.9	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1069	83.9	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1067	83.8	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1065	83.6	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5 Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5 Q7Z0W5	Q7z0w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	244.5	19.2	236	5	Q8T6U0	Q8t6u0 dendroneph
19	242.5	19.0	225	5	Q7Z0W9	Q7z0w9 montastraea
20	240	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyll
21	238.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q7Z0W4	Q7z0w4 montastraea
23	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
24	232	18.2	225	5	Q8T5F1	Q8t5f1 montastraea
25	214	16.8	227	5	Q7Z0W6	Q7z0w6 montastraea
26	214	16.8	234	5	Q7Z0W7	Q7z0w7 montastraea
27	212.5	16.7	259	5	Q8MMA2	Q8mma2 agaricia fr
28	210	16.5	239	5	Q8MMA1	Q8mma1 agaricia ag
29	209	16.4	234	5	Q8T5F2	Q8t5f2 montastraea
30	209	16.4	234	5	Q8MU47	Q8mu47 montastraea
31	208.5	16.4	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	206	16.2	227	5	Q962P9	Q962p9 montastraea
33	206	16.2	227	5	Q7Z0W8	Q7z0w8 montastraea
34	205.5	16.1	232	5	Q9GPI5	Q9gpi5 anemonia su
35	204.5	16.1	221	5	Q86LV7	Q86lv7 meandrina m
36	204	16.0	221	5	Q95P04	Q95p04 gonopora t
37	203.5	16.0	214	5	Q86LV8	Q86lv8 meandrina m
38	203.5	16.0	238	5	Q9BLY9	Q9bly9 renilla mue
39	202	15.9	227	5	Q95V70	Q95vt0 montastraea
40	201.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
41	201.5	15.8	232	5	Q9GZ28	Q9gz28 anemonia su
42	199	15.6	235	5	Q8T5F0	Q8t5f0 scolymia cu
43	198.5	15.6	222	5	Q72168	Q72168 cerianthus
44	198.5	15.6	225	5	Q8T6T9	Q8t6t9 radianthus
45	197.5	15.5	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2 ID Q8GHE2 PRELIMINARY; PRT; 238 AA.
AC Q8GHE2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 2289GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM2289;
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RL Azomonas and Azotobacter."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324408; AAN86140.1; -;
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR0009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;

Query Match 97.2%; Score 1238; DB 2; Length 238;

Best Local Similarity 97.5%; Pred.No. 1.1e-95;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGVNGHKFSVSGEGDATYKLTIKFICTTGKLPVWPTL 61

Db 1 MSKGEELFTGVVPIVLVDGVNGHKFSVSGEGDATYKLTIKFICTTGKLPVWPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVFEQDTLV 121

Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181

Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMVLGAVTAAAGITLGMDELYK 239

Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMVLLEFVTAAGITHGMDELYK 238

RESULT 2

Q93125 PRELIMINARY; PRT; 238 AA.

AC Q93125

DT 01-FEB-1997 (TrEMBLrel. 02, Created)

DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Green fluorescent protein mutant 3.

GN GFP.

OS Aequorea victoria (Jellyfish).

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=6100;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96305137; PubMed=8707053;

RA Cormack B.P., Valdivia R.H., Falkow S.;

RT "FACS-optimized mutants of the green fluorescent protein (GFP).";

RL Gene 173:33-38(1996).

RN [2]

RP SEQUENCE FROM N.A.

RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,

RA Brown A.J.P.;

RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene

RT expression in Candida albicans.;"

RL Microbiology 0:0-0(1996).

DR EMBL; U73901; AAB18957.1; -.

DR HSSP; P42212; 1BFP.

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP_like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP; 1.

DR ProDom; PD013756; Green_fl_protein; 1.

SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 96.9%; Score 1235; DB 5; Length 238;

Best Local Similarity 97.1%; Pred. No. 1.9e-95;

Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFCITCTGKLPVWPPTL 61

Db 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFCITCTGKLPVWPPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 121

Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181

Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMVLGAVTAAAGITLGMDELYK 239

Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMVLLEFVTAAGITHGMDELYK 238

RESULT 3

Q8GHE4 PRELIMINARY; PRT; 238 AA.

AC Q8GHE4

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DE Green fluorescent protein.

GN 85GFP.

OS Azotobacter vinelandii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Azotobacter.

OX NCBI_TaxID=354;

RN [1]

RP SEQUENCE FROM N.A.

RA Koranyi P., Berenyi M., Burg K.;

RT "Occurrence of green fluorescence protein in diazotrophic bacteria

RT Azomonas and Azotobacter.;"

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF324406; AAN86138.1; -.

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP; 1.

DR ProDom; PD013756; Green_fl_protein; 1.

SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.7%; Score 1232; DB 2; Length 238;

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Green fluorescence protein.

GN 375GFP.

OS Azomonas agilis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Azomonas.

OX NCBI_TaxID=116849;

RN [1]

RP SEQUENCE FROM N.A.

RA Koranyi P., Berenyi M., Burg K.;

RT "Occurrence of green fluorescence protein in diazotrophic bacteria

RT Azomonas and Azotobacter.;"

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF324405; AAN86137.1; -.

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP; 1.

DR ProDom; PD013756; Green_fl_protein; 1.

SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.8%; Score 1233; DB 2; Length 238;

Best Local Similarity 97.1%; Pred. No. 2.9e-95;

Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFCITCTGKLPVWPPTL 61

Db 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFCITCTGKLPVWPPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 121

Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181

Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMVLGAVTAAAGITLGMDELYK 239

Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMVLLEFVTAAGITHGMDELYK 238

RESULT 4

Q8GHE3 PRELIMINARY; PRT; 238 AA.

AC Q8GHE3

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DE Green fluorescence protein.

GN 85GFP.

OS Azotobacter vinelandii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Azotobacter.

OX NCBI_TaxID=354;

RN [1]

RP SEQUENCE FROM N.A.

RA Koranyi P., Berenyi M., Burg K.;

RT "Occurrence of green fluorescence protein in diazotrophic bacteria

RT Azomonas and Azotobacter.;"

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF324406; AAN86138.1; -.

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP; 1.

DR ProDom; PD013756; Green_fl_protein; 1.

SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.7%; Score 1232; DB 2; Length 238;


```
Best Local Similarity 97.1%; Pred. No. 3.5e-95;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLKFCITCTGKLPVPWPTL 61
DB 1 MSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLKFCITCTGKLPVPWPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTRAEVKFEGLTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNLGHKLEYNHSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNLGHKLEYNHSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAVTAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR00017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.2%; Score 1200; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 1.7e-92;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLKFCITCTGKLPVPWPTL 61
DB 1 MSKGEELFTGVWPILVELDGVNGKFVSVEGEGDATYKGLTLKFCITCTGKLPVPWPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTRAEVKFEGLTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNLGHKLEYNHSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNLGHKLEYNHSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAVTAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
```

```
Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR00017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; BD4648262D8EABD4 CRC64;

Query Match 93.0%; Score 1185; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 3e-91;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLKFCITCTGKLPVPWPTL 61
DB 1 MSKGEELFTGVWPILVELDGVNGKFVSVEGEGDATYKGLTLKFCITCTGKLPVPWPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTRAEVKFEGLTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNLGHKLEYNHSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNLGHKLEYNHSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAVTAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; RAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR00017; GFP like.
DR Pfam; PF01353; GFP; 1.
```

DR	PRINTS; PRO1229; GFLUORESCENT.
DR	ProDom; PD013756; Green_fl_protein; 1.
SQ	SEQUENCE 238 AA; 27015 MW; 6B8FD75E88926903 CRC64;
Query Match 84.8%; Score 1080; DB 5; Length 238;	
Best Local Similarity 81.9%; Pred. No. 1.8e-82;	
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;	
QY	2 VSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTLCFTCTGKLPVWPPTL 61
DB	1 MSKGEELFTGIVPVLIELDGVHGHKFSVRGEGDADYGKLEIKFTCTGKLPVWPPTL 60
QY	62 VTTLISGVQCFRSYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB	61 VTTLISYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY	122 NRLEKGDIFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB	121 NRLEKGMDFKEDGNILGHKLEYNFNSHNVYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180
QY	182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGVTAAGITLGMDELYK 239
DB	181 HYQTNVPLGDGPVLIPINHVLTQTATSKDRNETRDHMFLEFFSACGHTGMDELYK 238
RESULT 8	
ID	Q8WP95 PRELIMINARY; PRT; 238 AA.
AC	Q8WP95;
DT	01-MAR-2002 (T-EMBLrel. 20, Created)
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE	Green fluorescent protein.
GN	GFPXM.
OS	Aequorea macrodactyla.
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC	Aequoreidae; Aequorea.
OX	NCBI_TaxID=147615;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=GFPxm161;
RA	Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qian Y.X., Pang S.Q.,
RA	Li S.J., Xia N.S.;
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AY013824; AAK02059.1; -.
DR	GO; GO:0006091; P:energy pathways; IEA.
DR	InterPro; IPR009017; GFP like.
DR	Pfam; PF01353; GFP; 1.
DR	PRINTS; PRO1229; GFLUORESCENT.
DR	ProDom; PD013756; Green_fl_protein; 1.
SQ	SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;
Query Match 84.5%; Score 1076; DB 5; Length 238;	
Best Local Similarity 81.5%; Pred. No. 4e-82;	
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;	
QY	2 VSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTLCFTCTGKLPVWPPTL 61
DB	1 MSKGEELFTGIVPVLIELDGVHGHKFSVRGEGDADYGKLEIKFTCTGKLPVWPPTL 60
QY	62 VTTLISGVQCFRSYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB	61 VTTLISYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY	122 NRLEKGDIFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB	121 NRLEKGMDFKEDGNILGHKLEYNFNSHNVYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180
QY	182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGVTAAGITLGMDELYK 239
DB	181 HYQTNVPLGDGPVLIPINHVLTQTATSKDRNETRDHMFLEFFSACGHTGMDELYK 238
RESULT 9	
ID	Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC	Q8WTC4;
DT	01-MAR-2002 (T-EMBLrel. 20, Created)
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE	Green fluorescent protein.
GN	GFP.
OS	Aequorea macrodactyla.
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC	Aequoreidae; Aequorea.
OX	NCBI_TaxID=147615;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Shg24;
RA	Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA	Li S.J., Xia N.S.;
RL	"Colorful mutants of green fluorescent protein from Aequorea
RL	macrodactyla."; 2001 to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF435433; AAL33918.1; -.
DR	GO; GO:0006091; P:energy pathways; IEA.
DR	InterPro; IPR009017; GFP like.
DR	Pfam; PF01353; GFP; 1.
DR	PRINTS; PRO1229; GFLUORESCENT.
DR	ProDom; PD013756; Green_fl_protein; 1.
SQ	SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;
Query Match 84.1%; Score 1072; DB 5; Length 238;	
Best Local Similarity 82.4%; Pred. No. 8.6e-82;	
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;	
QY	2 VSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTLCFTCTGKLPVWPPTL 61
DB	1 MSKGEELFTGVVPIVLVELDGVHGHKFSVRGEGDADYGKLEIKFTCTGKLPVWPPTL 60
QY	62 VTTLISGVQCFRSYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB	61 VTTLISYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY	122 NRLEKGDIFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB	121 NRLEKGMDFKEDGNILGHKLEYNFNSHNVYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180
QY	182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLAGVTAAGITLGMDELYK 239
DB	181 HYQTNVPLGDGPVLIPINHVLTQTATSKDRNETRDHMFLEFFSACGHTGMDELYK 238
RESULT 10	
ID	Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC	Q8WTD0;
DT	01-MAR-2002 (T-EMBLrel. 20, Created)
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE	Green fluorescent protein.
GN	GFP.
OS	Aequorea macrodactyla.
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC	Aequoreidae; Aequorea.
OX	NCBI_TaxID=147615;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=GFPxm161;
RA	Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA	Li S.J., Xia N.S.;
RL	"Colorful mutants of green fluorescent protein from Aequorea
RL	macrodactyla.";

```
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL33912.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match      84.0%; Score 1070; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.5e-81;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLFICTTGKLPVPWPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGIVPVLIELDGDVHGKFSVSGEGDADYKGLKLEIKTCTTGKLPVPWPTL 60

QY 62 VTLSYGVQCFSPYDPMKQHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTLGYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 NRLELKGMDFKEDGNILGHKLEYNHNHNYIMPDKANGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HYQTNVPLGDGPVLLPINHYLSQTAISKDRNETRDHVMVLEFFSACGHTGHMDELYK 238

RESULT 11
Q8WTC8 PRELIMINARY; PRT; 238 AA.
AC Q8WTC8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPxm163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL33913.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match      83.9%; Score 1069; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.5e-81;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLFICTTGKLPVPWPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGIVPVLIELDGDVHGKFSVSGEGDADYKGLKLEIKTCTTGKLPVPWPTL 60

QY 62 VTLSYGVQCFSPYDPMKQHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTLGYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 NRLELKGMDFKEDGNILGHKLEYNHNHNYIMPDKANGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HYQTNVPLGDGPVLLPINHYLSQTAISKDRNETRDHVMVLEFFSACGHTGHMDELYK 238

RESULT 13
Q8WTC7 PRELIMINARY; PRT; 238 AA.
AC Q8WTC7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435429; AAL33914.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match      83.9%; Score 1069; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.5e-81;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLFICTTGKLPVPWPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGIVPVLIELDGDVHGKFSVSGEGDADYKGLKLEIKTCTTGKLPVPWPTL 60

QY 62 VTLSYGVQCFSPYDPMKQHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTLGYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds
(without alignments)
1433.395 Million cell updates/sec

Title: US-09-887-784-222G

Perfect score: 1276

Sequence: 1 MYSKGEELFTGVVILVELD.....VLGGFVTRAGITLGMDELK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Genesecp1980s:*
- 2: Genesecp1990s:*
- 3: Genesecp2000s:*
- 4: Genesecp2001s:*
- 5: Genesecp2002s:*
- 6: Genesecp2003as:*
- 7: Genesecp2003bs:*
- 8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	99.2	239	5	AAE17518 Enhanced
2	1266	99.2	363	6	ABR40352 Human ami
3	1266	99.2	893	4	AAG65781 Amino aci
4	1266	99.2	1132	4	AAG55782 Amino aci
5	1258	98.6	239	5	AAE17517 Enhanced
6	1255	98.4	239	3	AAE17517 Enhanced
7	1255	98.4	239	3	AAE17517 Enhanced
8	1255	98.4	239	3	AAE17517 Enhanced
9	1255	98.4	239	4	AAE17517 Enhanced
10	1255	98.4	239	4	AAE17517 Enhanced
11	1255	98.4	239	4	AAE17517 Enhanced
12	1255	98.4	239	5	ABG66198 A. victor
13	1255	98.4	239	5	ABG66198 A. victor
14	1255	98.4	239	5	ABG66198 A. victor
15	1255	98.4	239	6	AAE14599 Aequorea
16	1255	98.4	239	6	AAE14599 Aequorea
17	1255	98.4	239	6	AAE14599 Aequorea
18	1255	98.4	239	6	AAE14599 Aequorea
19	1255	98.4	239	7	ABU63204 Aequorea
20	1255	98.4	239	7	ADCI18358 EGFP (enh
21	1255	98.4	239	7	ABW00914 Aequorea
22	1255	98.4	239	7	ABW00914 Aequorea
23	1255	98.4	246	7	ABW00914 Aequorea
24	1255	98.4	248	5	AAG68319 Jellyfish
25	1255	98.4	259	5	AAU99804 Biomembra

26	1255	98.4	265	2	AAW97451 Wild-type
27	1255	98.4	268	5	AAU99803 Biomembra
28	1255	98.4	270	5	AAU99803 Biomembra
29	1255	98.4	272	5	AAU99800 Biomembra
30	1255	98.4	273	5	AAU99801 Biomembra
31	1255	98.4	280	5	AAU99807 Biomembra
32	1255	98.4	281	3	AAU50142 Green flu
33	1255	98.4	281	3	AAU50142 Green flu
34	1255	98.4	281	5	AAU10888 EGFP-MODC
35	1255	98.4	286	7	AAU10888 EGFP-MODC
36	1255	98.4	289	7	AAU10888 EGFP-MODC
37	1255	98.4	290	7	AAU10888 EGFP-MODC
38	1255	98.4	294	3	AAU10888 EGFP-MODC
39	1255	98.4	294	3	AAU10888 EGFP-MODC
40	1255	98.4	294	3	AAU10888 EGFP-MODC
41	1255	98.4	294	5	ABG94422 Recombina
42	1255	98.4	308	2	AAU42181 EGFP/DRM
43	1255	98.4	320	6	ABR83620 HUB1-GFP
44	1255	98.4	323	3	AAU54359 GFP mutan
45	1255	98.4	323	6	ABR83621 RUB1-GFP

ALIGNMENTS

RESULT 1

AAE17518

ID AAE17518 standard; protein; 239 AA.

AC AAE17518;

DT 22-APR-2002 (first entry)

DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
cellular function; genetic reporter; mutant; Stoke's shift; mutein.

OS Aequorea victoria.

OS Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 65

FT /note= "Wild type Phe substituted with Leu; This
corresponds to position 64 in the wild type protein"

FT Misc-difference 223

FT /note= "Wild type Glu substituted with Gly; This
corresponds to position 222 in the wild type protein"

FT WO200198338-A2.

PD 27-DEC-2001.

PF 18-JUN-2001; 2001WO-EP006848.

PR 19-JUN-2000; 2000DK-00000953.

PR 20-JUN-2000; 2000US-0212681P.

PR 10-MAY-2001; 2001DK-00000739.

PR 10-MAY-2001; 2001US-0290170P.

XX (BIOI-) BIOIMAGE AS.

XX Bjorn SP, Pagliaro L, Thastrup O;

XX WPI; 2002-098224/13.

DR N-PSDB; AAD28163.

XX Novel fluorescent protein in vitro assay for measuring protein kinase

PT activity or dephosphorylation activity, or for measuring protein

PT redistribution, has a green fluorescent protein with F64L and E222G

PT mutation.

XX Claim 9; Page 37; 4lpp; English.

PS

XX The invention relates to a fluorescent protein derived from green
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
 CC at F64L and E222G has a bigger compared to other GFP's making it very
 CC suitable for high throughput screening due to better resolution. The
 CC fluorescent protein is useful in vitro assays for measuring protein
 CC kinase activity or dephosphorylation activity, or for measuring cellular
 CC redistribution. The fluorescent protein is useful in studying cellular
 CC functions in living cells; as protein tags in transgenic animals, living
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.
 CC The fluorescent protein is also useful as a cell or organelle integrity
 CC marker, a marker for changes in cell morphology, as transfection marker,
 CC and as a marker to be used in combination with fluorescence activated
 CC cell sorting (FACS). The novel proteins can also be used as reporters to
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
 CC protein is also useful as markers in transcriptional and translational
 CC fusions for performing transposon vector mutagenesis and as a reporter
 CC for bacterial detection. Transposons encoding the fluorescent protein are
 CC useful for screening promoters and for tagging plasmids and chromosomes.
 CC The fluorescent protein engineered into the genome of a phage is useful
 CC for designing diagnostic tool. The present sequence is a DNA encoding
 CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
 XX
 SQ Sequence 239 AA;

Query Match 99.2%; Score 1266; DB 5; Length 239;
 Best Local Similarity 99.8%; Pred. No. 1.4e-122;
 Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVWPWT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVWPWT 60
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239
 DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239

RESULT 2
 ABR40352
 ID ABR40352 standard; protein; 363 AA.

XX ABR40352;

XX 08-JUL-2003 (first entry)

XX Human amino acid sequence SEQ ID NO: 6.

XX Human; heterologous conjugate; intracellular protein.

XX Homo sapiens.
 OS Aequoria victoria.

XX WO2003029827-A2.

XX 10-APR-2003.

XX 01-OCT-2002; 2002WO-DK000651.

XX 01-OCT-2001; 2001DK-00001433.

XX 11-OCT-2001; 2001US-0328996P.

XX (BIOL-) BIOMAGE AS.

XX Terry BR, Nielsen SJ;

XX WPI: 2003-430211/40.
 DR N-PSDB; ACC72604.

XX Novel cell for identifying modulators of protein interaction, contains a
 PT first conjugate comprising anchor protein, second conjugate having type B
 PT interactor protein and third conjugate with detectable group.

XX Disclosure; Page 112-113; 118pp; English.

XX The invention relates to a novel cell, comprising three heterologous
 CC conjugates (HC), a first HC (HC1) comprising an anchor protein that
 CC specifically binds to an internal structure within the cell conjugated to
 CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
 CC type B conjugated to a first protein of interest, and a third HC (HC3)
 CC comprising a second protein of interest conjugated to detectable group.
 CC The cell is useful for detecting if a compound disrupts or induces the
 CC interaction between two intracellular proteins. The cell is also useful
 CC for screening compounds that modulate the interaction between two
 CC intracellular proteins. The present sequence is used in the
 CC exemplification of the invention

XX Sequence 363 AA;

Query Match 99.2%; Score 1266; DB 6; Length 363;
 Best Local Similarity 99.6%; Pred. No. 2.6e-122;
 Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVWPWT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVWPWT 60

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239
 DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239

RESULT 3
 AAG65781
 ID AAG65781 standard; protein; 893 AA.

XX AAG65781;

XX 07-JAN-2002 (first entry)

XX Amino acid sequence of HSPDE4A1-E222G fusion protein.

XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
 KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
 XX fusion protein.

OS Homo sapiens.
 OS Aequorea victoria.

XX WO200179526-A2.

XX 25-OCT-2001.

XX 11-APR-2001; 2001WO-DK000264.

XX 17-APR-2000; 2000DK-00000651.

XX 29-MAY-2000; 2000DK-00000849.

XX (BIOI-) BIOIMAGE AS.

KW	Jellyfish; green fluorescent protein; GFP; protein redistribution; cellular function; genetic reporter; mutant; Stoke's shift; mitein.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 65	/note= "Wild type Phe substituted with Leu; This corresponds to position 64 in the wild type protein"
FT		
FT		
FT		
PN	WO200198338-A2.	
XX		
XX	27-DEC-2001.	
XX	18-JUN-2001; 2001WO-EP006848.	
XX	19-JUN-2000; 2000DK-00000953.	
PR	20-JUN-2000; 2000US-0212681P.	
PR	10-MAY-2001; 2001DK-00000739.	
PR	10-MAY-2001; 2001US-0290170P.	
XX		
PA	(BIOI-) BIOIMAGE AS.	
XX		
PI	Bjorn SP, Pagliaro L, Thastrup O;	
XX		
DR	WPI; 2002-098224/13.	
DR	N-PSDE; AAD28162.	
XX		
XX	Novel fluorescent protein in in vitro assay for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution, has a green fluorescent protein with F64L and E222G mutation.	
XX		
PS	Example 1; Page 35; 41pp; English.	
XX		
CC	The invention relates to a fluorescent protein derived from green fluorescent protein (GFP) or its analogue. The GFP containing mutations at F64L and E222G has a bigger compared to other GFP's making it very suitable for high throughput screening due to better resolution. The fluorescent protein is useful in invitro assays for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution. The fluorescent protein is useful in studying cellular functions in living cells; as protein tags in transgenic animals, living and fixed cells; organelle tags, secretion marker and genetic reporter. The fluorescent protein is also useful as a cell or organelle integrity marker, a marker for changes in cell morphology, as transfection marker, and as a marker to be used in combination with fluorescence activated cell sorting (FACS). The novel proteins can also be used as reporters to monitor live or dead biomass of organisms, such as fungi. The fluorescent protein is also useful as markers in transcriptional and translational fusions for performing transposon vector mutagenesis and as a reporter useful for bacterial detection. Transposons encoding the fluorescent protein are useful for screening promoters and for tagging plasmids and chromosomes. The fluorescent protein engineered into the genome of a phage is useful for designing diagnostic tool. The present sequence is enhanced F64L jellyfish green fluorescent protein (GFP) mutant	
XX		
SQ	Sequence 239 AA;	
	Query Match	98.6%; Score 1258; DB 5; Length 239;
	Best Local Similarity	99.2%; Pred. No. 9.4e-122;
	Matches 237; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 MVSGBELFTQVPIVLVDGVDVNGHKFSVSGEGGATYKLTLPKICTTGKLPVWPPT 60	
Db	1 MVSGBELFTQVPIVLVDGVDVNGHKFSVSGEGGATYKLTLPKICTTGKLPVWPPT 60	
Qy	61 LVTTLSVGVCFSYPPDHMKQHDFFKSNMPGGYVQERTIFFKDDGNKYKTRAEVKFEGDTL 120	
Db	61 LVTTLSVGVCFSYPPDHMKQHDFFKSNMPGGYVQERTIFFKDDGNKYKTRAEVKFEGDTL 120	
Qy	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQKNGIKVNFKIRHNIEDGSVQLA 180	

CC biosensors of the invention can be used to investigate a wide range of
CC cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased quantity
CC and speed of data collection and faster evaluation of drug candidates.
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used
CC as components of biosensor fusion proteins of the invention
XX
SQ Sequence 239 AA;

Query Match 98.4%; Score 1255; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MYSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTLFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTLFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

RESULT 7
AAV54349
ID AAY54349 standard; protein; 239 AA.
XX AAY54349;
XX AC AAY54349;
XX DT 06-APR-2000 (first entry)
DE Amino acid sequence of the mutant green fluorescent protein EGFP.
XX Fluorescent protein; green fluorescent protein; emission intensity;
KW fluorescence; pH detection; pH sensor; EGFP.
XX Synthetic.
OS Aequorea victoria.

Key Location/Qualifiers
FH Misc-difference 65
FT Misc-difference 66 /note= "wild type Phe substituted with Leu"
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"
FT Misc-difference 232 /note= "wild type His substituted with Leu"
XX WO9964592-A2.
XX 16-DEC-1999.
XX 08-JUN-1999; 99WO-US012850.
XX 09-JUN-1998; 98US-00094359.
XX 13-OCT-1998; 98US-00172063.
XX (REGC) UNIV CALIFORNIA.
XX (UYOR-) UNIV OREGON STATE.

PI Tsien RY, Llopis J, Wachtter RM;
XX WPI; 2000-116540/10.
DR N-PSDB; AA245642.
XX New functional engineered green fluorescent proteins, used for measuring
PT the pH in biological samples and cells.
PT Disclosure; Page 9; 89pp; English.
PS
XX The present sequence represents a functional engineered fluorescent
CC protein based on the Aequorea green fluorescent protein (GFP). The
CC emission intensity changes as pH varies between 5 and 10 of the present
CC protein are novel. The functional engineered fluorescent proteins show
CC reversible changes in fluorescence over physiological pH ranges. They can
CC be used for determining the pH of samples and cells. The polynucleotides
CC can also be used to produce transgenic animals. The fluorescent protein
CC pH sensors can be delivered to cells in the form of polynucleotides
CC encoding the protein sensor fused to a targeting signal. The targeting
CC signal directs the expression of the protein sensors to restricted cell
CC locations. This makes it possible to measure the pH of a precisely
CC defined cellular region or organelle
XX
SQ Sequence 239 AA;

Query Match 98.4%; Score 1255; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MYSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTLFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLTLFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

RESULT 8
AAV79584
ID AAY79584 standard; peptide; 239 AA.
XX AAY79584;
XX AC AAY79584;
XX DT 29-AUG-2000 (first entry)
XX EGFP signal domain.
DE EGFP signal domain.
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;
KW analysis; drug discovery.
XX Unidentified.
OS WO200026408-A2.
XX 11-MAY-2000.
XX 29-OCT-1999; 99WO-US025431.
XX 30-OCT-1998; 98US-0106308P.
XX 26-MAY-1999; 99US-0136078P.
XX (CELL-) CELLOMICS INC.
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;


```

PD XX 20-SEP-2001.
PF XX
PR XX 14-MAR-2001; 2001WO-US008071.
PR XX
PR XX 15-MAR-2000; 2000US-0189698P.
PR XX
PA XX (MERI ) MERCK & CO INC.
PA XX
PI XX Marsh DJ;
PI XX
XX XX
DR XX WPI; 2001-565791/63.
DR XX N-PSDB; AAH47304.
XX XX
XX XX Fusion proteins comprising melanin concentrating hormone receptor
PT XX peptides and fluorescent proteins, useful for identifying appetite
PT XX stimulants.
XX XX
PS Claim 2; Page 14; 7lpp; English.
XX XX
CC The invention provides melanin concentrating hormone (MCH) receptor
CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
CC MCHR polypeptide regions from different species. The MCHR fusion protein
CC comprise MCHR polypeptide region and a fluorescent polypeptide region
CC joined directly, or via a linker, to the carboxy side of the MCHR
CC polypeptide region. The MCHR fusion proteins can be expressed by standard
CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
CC regulation of MCH activity stimulates food intake. The present sequence
CC represents a A. victoria green fluorescent protein (GFP) and a linker
CC sequence
XX XX
SQ Sequence 239 AA;

Query Match 98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEEFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MYSKGEEFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 11
AAB31171
ID AAB31171 standard; protein; 239 AA.
XX
AC AAB31171;
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a green fluorescent protein (GFP).
XX
XX Growth rate; death rate; reporter gene; luminescent protein;
KW fluorescent product; luciferase; green fluorescent protein; GFP.
XX
XX Aequorea victoria.
OS
XX WO200075367-A1.
PN
XX 14-DEC-2000.
PD
XX 07-JUN-2000; 2000WO-FI000507.
PF

XX 07-JUN-1999; 99FI-00001296.
XX
XX (LILI/) LILIUS E.
PA (VIRT/) VIRT M.
XX
XX Lilius E, Virta M;
PI
XX WPI; 2001-061737/07.
DR N-PSDB; AAC86954.
XX
XX Assessing growth and death rates of a micro-organism in a desired
PT environment, by introducing 2 reporter genes encoding luminescent and
PT fluorescent products and detecting luminescent fluorescence.
XX
PS Disclosure; Page 27; 32pp; English.
XX
XX The specification describes a method for assessing the growth rate and
CC death rate of a micro-organism within a predetermined time period in a
CC desired environment. The method comprises introducing at least two
CC reporter genes encoding luminescent and/or fluorescent products into the
CC micro-organisms, incubating the micro-organism within the desired a
CC environment, and detecting luminescence and/or fluorescence after a
CC predetermined time period. Use of two different markers within a micro-
CC organism enables the differentiation between growth and death rates. The
CC method is used to assess the growth rate and death rate of a micro-
CC organism within a predetermined time period in a desired environment. The
CC present sequence represents a green fluorescent protein (GFP), and is
CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
CC and is used in the method of the invention
XX
XX Sequence 239 AA;

Query Match 98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEEFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MYSKGEEFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 12
AAG66198
ID AAG66198 standard; protein; 239 AA.
XX
AC AAG66198;
XX
XX 17-JUN-2002 (first entry)
XX
XX A. victoria green fluorescent protein (EGFP).
XX
XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
KW green fluorescent protein; EGFP.
XX
XX Aequorea victoria.
OS
XX JP2002045189-A.
PN
XX 12-FEB-2002.
PD
XX

```

PF	04-AUG-2000; 2000JP-00237165.	
XX	PR	
PR	04-AUG-2000; 2000JP-00237165.	
XX	(RIKA) RIRAGAKU KENKYUSHO.	
XX	PA	
XX	WPI; 2002-299190/34.	
DR	N-PSDB; ABL40628.	
XX		
PT	A gene encoding cyan-green fluorescent protein.	
XX		
PS	Example; Page 14; 20pp; Japanese.	
XX		
CC	The invention relates to a gene encoding proteins having cyan-green	
CC	fluorescence characteristic and having a function of showing stable	
CC	fluorescence characteristic in acid region. A method for the preparation	
CC	of a cyan-green fluorescent protein is provided which involves a	
CC	transformant transformed by a recombinant vector comprising the gene,	
CC	where the transformant is cultured and the protein is collected from the	
CC	culture. The present sequence represents the A. victoria green	
CC	fluorescent protein (EGFP)	
XX		
SQ	Sequence 239 AA;	
	Query Match 98.4%; Score 1255; DB 5; Length 239;	
	Best Local Similarity 98.7%; Pred. No. 1.9e-121;	
	Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
QY	1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60	
DB	1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60	
QY	61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120	
DB	61 LVTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120	
QY	121 VNRLELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180	
DB	121 VNRLELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180	
QY	181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVGGFVTAAGITLGMDELYK 239	
DB	181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVGGFVTAAGITLGMDELYK 239	
RESULT 13		
ABG9444		
ID	ABG94444 standard; protein; 239 AA.	
XX		
AC	ABG94444;	
XX		
DT	27-NOV-2002 (first entry)	
XX		
DE	Protease biosensor signal sequence #6.	
XX		
XX	Detection; classification; identification; toxin detection; protease;	
KW	ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;	
KW	toxic threat agent.	
XX		
OS	Synthetic.	
XX		
PN	US6416959-B1.	
XX		
PD	09-JUN-2002.	
XX		
XX	25-FEB-2000; 2000US-00513783.	
XX		
PR	27-FEB-1997; 97US-00810983.	
PR	27-FEB-1998; 98US-00031271.	
PR	26-FEB-1999; 99US-0122152P.	
PR	08-MAR-1999; 99US-0123399P.	
PR	12-JUL-1999; 99US-00352171.	
PR	31-AUG-1999; 99US-0151797P.	

PR	17-SEP-1999; 99US-00398965.	
PR	29-OCT-1999; 99US-00430656.	
PR	01-DEC-1999; 99US-0168408P.	
XX		
PA	(GIUL/) GIULIANO K.	
PA	(KAPU/) KAPUR R.	
XX		
PI	Giuliano K, Kapur R;	
XX		
DR	WPI; 2002-634730/68.	
DR	N-PSDB; ABS71491.	
XX		
PT	Automated cell-based toxin detection, classification, and/or	
PT	identification by treating cells involves use of three classes of	
PT	luminescent reporter molecules such as detectors, classifiers or	
PT	identifiers.	
XX		
PS	Example 10; Fig 29A; 214pp; English.	
XX		
CC	The invention describes methods of automated detection, classification	
CC	and identification comprising treating cells containing luminescent	
CC	reporter molecules (1) in array of locations with a test substance, where	
CC	(1) are detectors, classifiers or identifiers, imaging cells in each	
CC	location to obtain luminescent signals and converting optical information	
CC	into digital data to interpret presence of toxins in the test substance.	
CC	The method are useful for detection of toxins chosen from proteases, ADP-	
CC	ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.	
CC	Three classes of cell-based luminescent reporter molecules such as	
CC	detectors, classifiers and identifiers are described and serve as	
CC	reporters of toxic threat agents. The first two levels of	
CC	characterisation ensure a rapid readout of toxin class without	
CC	sacrificing the ability to detect many new mutant toxins or dissect	
CC	several complex mixtures of known toxins. This is the amino acid sequence	
CC	of a protease biosensor related signal sequence used in the cell-based	
CC	screening system	
XX		
SQ	Sequence 239 AA;	
	Query Match 98.4%; Score 1255; DB 5; Length 239;	
	Best Local Similarity 98.7%; Pred. No. 1.9e-121;	
	Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
QY	1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60	
DB	1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60	
QY	61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120	
DB	61 LVTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120	
QY	121 VNRLELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180	
DB	121 VNRLELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180	
QY	181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVGGFVTAAGITLGMDELYK 239	
DB	181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVGGFVTAAGITLGMDELYK 239	
RESULT 14		
AAE14599		
ID	AAE14599 standard; protein; 239 AA.	
XX		
AC	AAE14599;	
XX		
DT	31-MAY-2002 (first entry)	
XX		
DE	Aequorea victoria enhanced green fluorescent protein.	
XX		
KW	Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; mutein.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	

XX FH Key Location/Qualifiers
 FT Misc-difference 1. .3
 FT Ser"
 FT Misc-difference 65
 FT /note= "GFP Phe64 is replaced by Leu"
 FT Misc-difference 66
 FT /note= "GFP Ser65 is replaced by Thr"
 XX EP1178109-A1.
 PN 06-FEB-2002.
 PD 03-AUG-2001; 2001EP-00306650.
 XX 04-AUG-2000; 2000JP-00237166.
 PR (RIKE) RIKEN KK.
 XX Miyawaki A, Sawano A;
 PI WPI: 2002-208112/27.
 XX N-PSDB; AAD27910.
 DR Method for mutagenesis, e.g. for introducing certain or random mutations
 XX at certain sites of the nucleotide sequence, comprises synthesizing a
 PT mutated strand and a complementary strand by use of megaprimers.
 PT Example 1; Page 13-14; 31pp; English.
 PS The invention relates to a method for mutagenesis that comprises
 XX synthesising a mutated strand and a complementary strand by use of
 CC megaprimers. The method basically comprises a DNA synthesis in which one
 CC or more primers that have a nucleotide sequence containing at least one
 CC mutation and a phosphorylated 5'-terminus are annealed to a template DNA
 CC and then subjected to an elongation reaction using a thermostable high-
 CC fidelity DNA polymerase, after which the phosphorylated 5'-terminus and
 CC the elongated terminus are ligated by means of a thermostable DNA ligase
 CC to synthesise a circular DNA containing the primers; a digestion in which
 CC the step of DNA synthesis is repeated several times to amplify the DNA
 CC containing the primers and then, at least DNAs other than the amplified
 CC circular DNA are digested into several fragments; and a double-stranded
 CC DNA synthesis in which, with the several fragments obtained in the step
 CC of digestion as megaprimers, the megaprimers are annealed to the circular
 CC DNA synthesised above, followed by an elongation reaction performed using
 CC the thermostable high-fidelity DNA polymerase. The method is useful for
 CC mutagenesis, particularly for introducing certain mutations at certain
 CC sites of a nucleotide sequence, or for introducing random mutations at
 CC certain sites of the nucleotide sequence. The present method is simple,
 CC speedy, economical and widely applicable. The present sequence is
 CC Aequorea victoria enhanced green fluorescent protein (EGFP) used for
 CC mutagenesis in an exemplification of the invention. The EGFP is derived
 CC by altering the green fluorescent protein (GFP) sequence of Aequorea
 CC victoria
 XX Sequence 239 AA;
 SQ
 Query Match 98.4%; Score 1255; DB 5; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.9e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTAKFICTTGKLPVPWPT 60
 DB 1 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTAKFICTTGKLPVPWPT 60
 QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
 RESULT 15
 AAE34958
 ID AAE34958 standard; protein; 239 AA.
 XX AAE34958;
 AC AAE34958;
 XX 28-MAY-2003 (first entry)
 DT Aequorea victoria enhanced green fluorescent protein (EGFP).
 DE Phosphorylation indicator; fluorescent protein; detection; phosphatase;
 XX Kinase; enhanced green fluorescent protein; EGFP.
 KW Aequorea victoria.
 OS WO200295058-A2.
 PN 28-NOV-2002.
 XX 24-MAY-2002; 2002WO-US016955.
 PF 24-MAY-2001; 2001US-00865291.
 PR (REGC) UNIV CALIFORNIA.
 XX Tsien RY, Ting AY, Zhang J;
 PI WPI: 2003-148474/14.
 DR N-PSDB; AAD53428.
 XX Novel chimeric phosphorylation indicators, useful for detecting
 PT kinase/phosphatase in samples, has donor molecule, phosphorylatable
 PT domain, phosphoaminoacid binding domain, and acceptor molecule, in
 PT operative linkage.
 XX Disclosure; Col 56-57; 38pp; English.
 PS The present invention relates to chimeric phosphorylation indicators
 CC comprising a phosphorylation polypeptide and a fluorescent protein or in
 CC operative linkage, a donor molecule, a phosphorylatable domain, a
 CC phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The
 CC phosphorylation indicators of the invention are useful for detecting
 CC kinases or phosphatases in a biological sample. They are also useful in
 CC high throughput analysis e.g. for detecting a kinase inhibitor or
 CC phosphatase inhibitor. The present sequence is Aequorea victoria enhanced
 CC green fluorescent protein (EGFP) used in the invention
 XX Sequence 239 AA;
 SQ
 Query Match 98.4%; Score 1255; DB 6; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.9e-121;
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTAKFICTTGKLPVPWPT 60
 DB 1 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTAKFICTTGKLPVPWPT 60
 QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

Db 131 DHYQNTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:13
Job time : 47.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-222G

Perfect score: 1276

Sequence: 1 MVSKGELFTGVVPIVELD.....VLGGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents_AA:*
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1255	98.4	239	3	US-09-172-063-3
2	1255	98.4	239	4	US-09-513-783A-46
3	1255	98.4	239	4	US-09-316-919-4
4	1255	98.4	239	4	US-09-602-641-3
5	1255	98.4	239	4	US-09-920-922-2
6	1255	98.4	281	3	US-09-062-102-1
7	1255	98.4	281	4	US-09-364-946-1
8	1255	98.4	294	4	US-09-513-783A-2
9	1255	98.4	323	3	US-09-172-063-21
10	1255	98.4	323	4	US-09-602-641-21
11	1255	98.4	364	3	US-09-085-305-6
12	1255	98.4	379	4	US-09-417-197-129
13	1255	98.4	434	4	US-09-800-170-48
14	1255	98.4	442	4	US-09-417-197-127
15	1255	98.4	459	4	US-09-513-783A-170
16	1255	98.4	544	4	US-09-417-197-113
17	1255	98.4	544	4	US-09-417-197-115
18	1255	98.4	604	4	US-09-417-197-59
19	1255	98.4	605	4	US-09-417-197-41
20	1255	98.4	606	4	US-09-417-197-65
21	1255	98.4	607	4	US-09-417-197-47
22	1255	98.4	630	4	US-09-417-197-63
23	1255	98.4	631	4	US-09-417-197-39
24	1255	98.4	633	4	US-09-417-197-45
25	1255	98.4	635	4	US-09-417-197-125
26	1255	98.4	642	2	US-08-818-253-2
27	1255	98.4	642	2	US-08-818-253-6

28	1255	98.4	642	3	US-08-818-252-2	Sequence 2, Appli
29	1255	98.4	642	3	US-08-818-252-6	Sequence 6, Appli
30	1255	98.4	652	3	US-08-818-253-4	Sequence 4, Appli
31	1255	98.4	652	3	US-08-818-252-4	Sequence 4, Appli
32	1255	98.4	718	4	US-09-417-197-75	Sequence 75, Appli
33	1255	98.4	719	4	US-09-417-197-51	Sequence 51, Appli
34	1255	98.4	726	4	US-09-417-197-71	Sequence 71, Appli
35	1255	98.4	727	4	US-09-417-197-139	Sequence 139, App
36	1255	98.4	783	4	US-09-513-783A-176	Sequence 176, App
37	1255	98.4	797	4	US-09-417-197-141	Sequence 141, App
38	1255	98.4	797	4	US-09-417-197-143	Sequence 143, App
39	1255	98.4	798	4	US-09-417-197-77	Sequence 77, Appli
40	1255	98.4	805	4	US-09-513-783A-178	Sequence 178, App
41	1255	98.4	806	4	US-09-417-197-53	Sequence 53, Appli
42	1255	98.4	836	4	US-09-417-197-61	Sequence 61, Appli
43	1255	98.4	842	4	US-09-417-197-43	Sequence 43, Appli
44	1255	98.4	843	4	US-09-417-197-117	Sequence 117, App
45	1255	98.4	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1
US-09-172-063-3
; Sequence 3, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Liopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-172-063-3

Query Match 98.4%; Score 1255; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKGELFTGVVPIVELDGVNKHKFSVSGEGDATYVKLTLLKFICTTGKLPVDPWT 60
DB 1 MVSKGELFTGVVPIVELDGVNKHKFSVSGEGDATYVKLTLLKFICTTGKLPVDPWT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNKYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIIDFKEDGNILGHKLEYNYNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180
DB 121 VNRIELKGIIDFKEDGNILGHKLEYNYNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLYSTQSALSCKDPNEKRDHNVLGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLYSTQSALSCKDPNEKRDHNVLGGFVTAAGITLGMDELYK 239

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Ilopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```

```
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Ilopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```



```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; CURRENT FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
US-09-062-102-1

Query Match      98.4%; Score 1255; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 3.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
US-09-364-946-1

Query Match      98.4%; Score 1255; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 3.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match      98.4%; Score 1255; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 3.3e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
|||||
```

Db 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTIFCTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVLGGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVLGGFVTAAGITLGMDELYK 239

RESULT 9
US-09-172-063-21
; Sequence 21, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-172-063-21

Query Match 98.4%; Score 1255; DB 3; Length 323;
Best Local Similarity 98.7%; Pred. No. 3.8e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTIFCTTGKLPVPWPT 60
Db 85 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTIFCTTGKLPVPWPT 144
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 204
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 205 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVLGGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVLGGFVTAAGITLGMDELYK 323

RESULT 10
US-09-602-641-21
; Sequence 21, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-602-641-21

Query Match 98.4%; Score 1255; DB 4; Length 323;
Best Local Similarity 98.7%; Pred. No. 3.8e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTIFCTTGKLPVPWPT 60
Db 85 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTIFCTTGKLPVPWPT 144
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 204
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 205 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVLGGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVLGGFVTAAGITLGMDELYK 323

RESULT 11
US-09-085-305-6
; Sequence 6, Application US/09085305
; Patent No. 6191269
; GENERAL INFORMATION:
; APPLICANT: Pollock, Allan
; APPLICANT: Lovett, David H.
; APPLICANT: Turck, Johanna
; TITLE OF INVENTION: Selective Induction of Apoptosis in
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,305
; FILING DATE: 29-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Francis, Carol L
REGISTRATION NUMBER: 36,513
REFERENCE/DOCKET NUMBER: 6510/102US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-085-305-6

Query Match 98.4%; Score 1255; DB 3; Length 364;
Best Local Similarity 98.7%; Pred. No. 4.6e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLTKFICTTGLKLPVPWPT 60
DB 126 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLTKFICTTGLKLPVPWPT 185

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 186 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRLEKIGDIFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 246 VNRLEKIGDIFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 305

QY 181 DHYQQNTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 306 DHYQQNTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 364

RESULT 12
US-09-417-197-129
Sequence 129, Application US/09417197
Patent No. 6518021
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE REFERENCE: 3759-0110P
CURRENT APPLICATION NUMBER: US/09/417,197
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: Patentin version 3.0
SEQ ID NO 129
LENGTH: 379
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: actin-binding-domain-EGFP fusion
US-09-417-197-129

Query Match 98.4%; Score 1255; DB 4; Length 379;
Best Local Similarity 98.7%; Pred. No. 4.9e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLTKFICTTGLKLPVPWPT 60
DB 141 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLTKFICTTGLKLPVPWPT 200

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 201 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRLEKIGDIFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 261 VNRLEKIGDIFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 320

QY 181 DHYQQNTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 321 DHYQQNTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 379

RESULT 13
US-09-800-170-48
Sequence 48, Application US/09800170
Patent No. 6481667
GENERAL INFORMATION:
APPLICANT: Kinsella, Todd
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
FILE REFERENCE: A-68614-1/DJB/RMS/RMK
CURRENT APPLICATION NUMBER: US/09/800,170
CURRENT FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: US 60/187,130
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 90
SOFTWARE: Patentin version 3.1
SEQ ID NO 48
LENGTH: 434
TYPE: PRT
ORGANISM: Synechocystis PCC6803
US-09-800-170-48

Query Match 98.4%; Score 1255; DB 4; Length 434;
Best Local Similarity 98.7%; Pred. No. 6e-127; 2; Indels 0; Gaps 0;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLTKFICTTGLKLPVPWPT 60
DB 196 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLTKFICTTGLKLPVPWPT 255

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 256 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRLEKIGDIFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 316 VNRLEKIGDIFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 375

QY 181 DHYQQNTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 376 DHYQQNTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 434

RESULT 14
US-09-417-197-127
Sequence 127, Application US/09417197
Patent No. 6518021
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE REFERENCE: 3759-0110P
CURRENT APPLICATION NUMBER: US/09/417,197
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: Patentin version 3.0
SEQ ID NO 127
LENGTH: 442
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EGFP-RhoA fusion
US-09-417-197-127

Query Match 98.4%; Score 1255; DB 4; Length 442;
Best Local Similarity 98.7%; Pred. No. 6.1e-127; 2; Indels 0; Gaps 0;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLTKFICTTGLKLPVPWPT 60

Db 1 MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFRSYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTGVQCFRSYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239

RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPP-HSP27
US-09-513-783A-170

Query Match 98.4%; Score 1255; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 6.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLPKFICTTGKLPVPWPT 60
Db 1 MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFRSYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTGVQCFRSYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:05
Job time : 12.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds
(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-222G
Perfect score: 1276
Sequence: 1 MVSKEELFTGVVPILVELD.....VLGGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
 - 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep:*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
 - 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
 - 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	
1	1266	99.2	239	9	US-09-887-784-4	Sequence 4, Appli
2	1266	99.2	239	12	US-10-296-953-4	Sequence 4, Appli
3	1266	99.2	363	14	US-10-270-223-6	Sequence 6, Appli
4	1266	99.2	893	14	US-10-257-909A-30	Sequence 30, Appli
5	1266	99.2	1132	14	US-10-257-909A-32	Sequence 32, Appli
6	1258	98.6	239	9	US-09-887-784-2	Sequence 2, Appli
7	1258	98.6	239	12	US-10-296-953-2	Sequence 2, Appli
8	1258	98.4	239	9	US-09-920-922-2	Sequence 2, Appli
9	1255	98.4	239	9	US-09-999-745-4	Sequence 4, Appli
10	1255	98.4	239	10	US-09-866-538-4	Sequence 4, Appli
11	1255	98.4	239	10	US-09-797-496B-2	Sequence 2, Appli
12	1255	98.4	239	10	US-09-794-308-4	Sequence 4, Appli
13	1255	98.4	239	10	US-09-865-291-4	Sequence 4, Appli
14	1255	98.4	239	12	US-10-457-982-3	Sequence 3, Appli
15	1255	98.4	239	14	US-10-121-258-13	Sequence 13, Appli

16	1255	98.4	239	14	US-10-221-461-7	Sequence 7, Appli
17	1255	98.4	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1255	98.4	239	14	US-10-177-390-2	Sequence 2, Appli
19	1255	98.4	239	14	US-10-338-411-3	Sequence 3, Appli
20	1255	98.4	239	15	US-10-370-570-4	Sequence 4, Appli
21	1255	98.4	239	15	US-10-389-640-3	Sequence 3, Appli
22	1255	98.4	259	14	US-10-314-861-11	Sequence 11, Appli
23	1255	98.4	281	12	US-09-931-232-1	Sequence 11, Appli
24	1255	98.4	288	14	US-10-314-861-37	Sequence 37, Appli
25	1255	98.4	293	14	US-10-314-861-35	Sequence 35, Appli
26	1255	98.4	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1255	98.4	295	14	US-10-314-861-39	Sequence 39, Appli
28	1255	98.4	299	14	US-10-314-861-33	Sequence 33, Appli
29	1255	98.4	305	14	US-10-314-861-31	Sequence 31, Appli
30	1255	98.4	308	14	US-10-033-717-35	Sequence 35, Appli
31	1255	98.4	311	14	US-10-314-861-29	Sequence 29, Appli
32	1255	98.4	320	14	US-10-338-411-11	Sequence 11, Appli
33	1255	98.4	320	15	US-10-389-640-11	Sequence 11, Appli
34	1255	98.4	323	12	US-10-457-982-21	Sequence 21, Appli
35	1255	98.4	323	14	US-10-338-411-7	Sequence 7, Appli
36	1255	98.4	323	14	US-10-338-411-13	Sequence 13, Appli
37	1255	98.4	323	15	US-10-389-640-7	Sequence 7, Appli
38	1255	98.4	323	15	US-10-389-640-13	Sequence 13, Appli
39	1255	98.4	324	14	US-10-314-861-16	Sequence 16, Appli
40	1255	98.4	345	14	US-10-338-411-5	Sequence 5, Appli
41	1255	98.4	345	15	US-10-389-640-5	Sequence 5, Appli
42	1255	98.4	346	14	US-10-338-411-9	Sequence 9, Appli
43	1255	98.4	346	15	US-10-389-640-9	Sequence 9, Appli
44	1255	98.4	359	14	US-10-033-717-33	Sequence 33, Appli
45	1255	98.4	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequoria Victoria
US-09-887-784-4

Query Match					99.2%;	Score 1266;	DB 9;	Length 239;
Best Local Similarity					99.6%;	Pred. No. 1.5e-123;		
Matches 238;					Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
QY	1	MVSKGEELFTGVVPILVELDGVN	GKFSVSGEGDATY	GKLT				
Db	1	MVSKGEELFTGVVPILVELDGVN	GKFSVSGEGDATY	GKLT				
QY	61	LVTTLSYGVCESRYPDHMKQHD	FKSAMPEGVQERTIFFKDDG	NYKTRAEVKFEGD	TL	120		
Db	61	LVTTLSYGVCESRYPDHMKQHD	FKSAMPEGVQERTIFFKDDG	NYKTRAEVKFEGD	TL	120		
QY	121	VNRIELKGIDFEDGNILGHKLE	YNSHNHYIMADKQNGIKVNF	KIRHNIEDSGVQ	LA	180		
Db	121	VNRIELKGIDFEDGNILGHKLE	YNSHNHYIMADKQNGIKVNF	KIRHNIEDSGVQ	LA	180		
QY	181	DHYQQNTPIGDGPVLLPDN	NHYLSTQSALSKDPNEKR	DHMLVGGFVTAAGIT	LGMDELYK	239		
Db	181	DHYQQNTPIGDGPVLLPDN	NHYLSTQSALSKDPNEKR	DHMLVGGFVTAAGIT	LGMDELYK	239		

RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: FAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match 99.2%; Score 1266; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.5e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

RESULT 3
US-10-270-223-6
; Sequence 6, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: Bioimage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACTING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION RE
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match 99.2%; Score 1266; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 2.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match 99.2%; Score 1266; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 9.8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match 99.2%; Score 1266; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 9.8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLKFICTTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLVGGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0

```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.2%; Score 1266; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 1.4e-122;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 953
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRLEKLGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVLGFGVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVLGFGVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJOERN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-887-784-2

Query Match          98.6%; Score 1258; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 1e-122;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVLGFGVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVLGFGVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
```

```
; APPLICANT: BJOERN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          98.6%; Score 1258; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 1e-122;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVLGFGVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVLGFGVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.4%; Score 1255; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
```


; APPLICANT: TSIEH, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match 98.4%; Score 1255; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
; Sequence 4, Application US/09865291
; Publication No. US20030186229A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEH, Roger
; APPLICANT: TING, Alice
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: EMISSION RATIO-METRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGEN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match 98.4%; Score 1255; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
RESULT 14
US-10-457-982-3
; Sequence 3, Application US/10457982
; Publication No. US20030212265A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/10/457,982
; PRIOR FILING DATE: 2003-06-09
; PRIOR APPLICATION NUMBER: US/09/602,641
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 98.4%; Score 1255; DB 12; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
RESULT 15
US-10-121-258-13
; Sequence 13, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CPI
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.4%; Score 1255; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 MVSKEBELFTGVVPILVELDGDVNGHKFVSYGEGGDATYGKLTAKFICTTGKLPVPWPT 60
Db      1 MVSKEBELFTGVVPILVELDGDVNGHKFVSYGEGGDATYGKLTAKFICTTGKLPVPWPT 60

QY      61 LVTTLSTGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKEGDTL 120
Db      61 LVTTLSTGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKEGDTL 120

QY      121 VNRLEKGIIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db      121 VNRLEKGIIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
Db      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:30
Job time : 34.7778 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds
(without alignments)
2224.817 Million cell updates/sec

Title: US-09-887-784-222G
Perfect score: 1276
Sequence: 1 MYSKGEELFTGVVILVELD.....VLGGFVTAAGITLGMDELYK 239
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1233	96.6	238	1 JQ1514	green-fluorescent
2	105	8.2	785	2 H72228	hypothetical prote
3	92	7.2	461	2 T06936	photosystem II chl
4	91.5	7.2	861	2 H64102	leucine-trna ligas
5	90	7.1	425	2 C97354	hypothetical prote
6	90	7.1	632	2 T06586	DNA-binding protei
7	89.5	7.0	887	2 E82590	leucyl-trna synthe
8	88.5	6.9	655	2 D83917	DNA topoisomerase
9	88	6.9	461	2 S41480	photosystem II chl
10	88	6.9	578	1 I40794	dihydrolipoamide d
11	87.5	6.9	370	2 E70390	iron-sulfur cofact
12	87.5	6.9	860	2 AC0582	leucyl-trna synthe
13	87.5	6.9	2222	1 A36028	DNA-directed DNA p
14	87.5	6.9	2573	2 D71614	hypothetical prote
15	87	6.8	1259	2 AB1055	probable exported
16	86	6.7	357	2 H81355	tRNA (uracil-5)-m
17	86	6.7	877	2 H64708	iron-regulated out
18	85.5	6.7	353	2 E84941	imidazolglycerol-
19	85.5	6.7	788	1 JQ1LHH	DNA-directed DNA p
20	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
21	85.5	6.7	1259	2 H65233	ytfn protein - Esc
22	85.5	6.7	1259	2 G91278	hypothetical prote
23	85.5	6.7	1259	2 G86119	hypothetical prote
24	85	6.7	281	2 AD2052	hypothetical prote
25	85	6.7	459	2 S42647	photosystem II chl
26	85	6.7	459	2 AD2342	photosystem II CP4
27	85	6.7	865	2 T11852	lipoygenase (EC 1
28	84.5	6.6	613	2 A99552	oligoendopeptidase
29	84.5	6.6	888	2 A54280	cell differentiat

glucose dehydrogen
synergohymenotropi
DNA-directed DNA p
Iga Fc receptor pr
Iga Fc receptor pr
hypothetical prote
transcription regu
photosystem II chl
S-layer protein pr
water-stress-induc
nitrogenase (EC 1.
neuraminidase, pro
protective surface
probable iron-regu
hypothetical prote

ALIGNMENTS

RESULT 1

JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C;Species: Aequorea victoria
C;Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C;Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331
R;Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A;Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A;Reference number: JQ1514; MUID:92175527; PMID:1347277
A;Accession: JS0692
A;Molecule type: DNA
A;Residues: 1-107, 'S', 109-238 <PRA1>
A;Cross-references: GB:M62654; NID:g155662; PIDN:AAA27722.1; PID:g155663
A;Accession: JQ1514
A;Molecule type: mRNA
A;Residues: 1-99, 'P', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A;Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661
A;Accession: PQ0335
A;Molecule type: protein
A;Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>
R;Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A;Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A;Reference number: S48693; MUID:94364470; PMID:8082767
A;Accession: S48693
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>
A;Cross-references: GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:g606384
R;Watkins, J.N.; Campbell, A.K.
Submitted to the EMBL data Library, January 1995
A;Reference number: S51330
A;Accession: S51330
A;Molecule type: mRNA
A;Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>
A;Cross-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009
A;Experimental source: clone gfp1
A;Accession: S51331
A;Molecule type: mRNA
A;Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 210-238 <PRA5>
A;Cross-references: EMBL:X83960; NID:g634010; PIDN:CAA58790.1; PID:g634011
A;Experimental source: clone gfp2
R;Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Submitted to the Brookhaven Protein Data Bank, August 1996
A;Reference number: A65692; PDB:IGFL
A;Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-91
A;Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
R;Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A;Title: The molecular structure of green fluorescent protein.
A;Reference number: A58953; MUID:98294543; PMID:9631087

A:Contents: annotation; X-ray crystallography, 1.9 angstroms
C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
C:Genetics:
A:Gene: GFP
A:Introns: 69/3; 167/3
C:Superfamily: green-fluorescent protein
C:Keywords: chromoprotein; luminescence
F;65-67/Cross-link: 5-midazolone (Ser-Gly) #status experimental
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 96.6%; Score 1233; DB 1; Length 238;
Best Local Similarity 96.6%; Pred. No. 9.1e-96;
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNNGHKFSVSGEGDATYKGLTKLFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGVNNGHKFSVSGEGDATYKGLTKLFICTTGKLPVWPPTL 60

QY 62 VTTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTFISYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDPKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVQLAD 181
DB 121 NRIELKGIDPKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNPKRDHMLGGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNPKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 2
H72228
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72228
R; Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: H72228
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-785 <ARN>
A:Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1624

Query Match 8.2%; Score 105; DB 2; Length 785;
Best Local Similarity 19.7%; Pred. No. 0.9;
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPILVELDGVNNGHKFSVSGEGDATYKGLTKLFICTTGKLPVWPPTLV 62
DB 15 NEGRFSGPVGTVGVQAD-----LVRKGLLPHYPVGM- 46

QY 63 TTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNKYKTRAEVKFEGDTLVN 122
DB 47 -----NEDLFKEIDREWIYEREFEFKEDVKGERVDLVFEGVDTLN 80

QY 123 RIELKGIDPKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVQLADH 182
DB 89 DVILNGVYL---GSTEDMFTEYRFDVTNLV-----KEKNLKVYIK-----SPIRVPKT 134

QY 183 YQONTPIGDGPVLLPDNHYLSQSALSKDPNPKRDHMLVGLGFTVTAAGITLGM 235
DB 135 LEQNYGLVGGP-----EDP-----IRGYIRKQAQSYGWD 163

RESULT 3
T09336
photosystem II chlorophyll a-binding protein psbC - Cyanophora paradoxa cyanelle
N:Alternate names: protein CP43
C:Species: cyanelle Cyanophora paradoxa
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 26-Aug-1999
C:Accession: T06936
R;Stewart, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A:Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A:Reference number: Z15840
A:Accession: T06936
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-461 <STI>
A:Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81279.1; PID:g1016192
A:Experimental source: strain Pringsheim LB555
C:Genetics:
A:Gene: psbC
A:Genome: cyanelle
C:Superfamily: photosystem II chlorophyll a-binding protein psbC
C:Keywords: chlorophyll; cyanelle; membrane-associated complex; photosynthesis; photosyst

Query Match 7.2%; Score 92; DB 2; Length 461;
Best Local Similarity 21.4%; Pred. No. 5.5;
Matches 58; Conservative 35; Mismatches 88; Indels 90; Gaps 13;

QY 25 GHKFSVSGEGDATYKGLTKLFICTTGK-----PVPWP--TLVTT-----LSYGV--- 69
DB 208 GDGWIVSDNMEDIIGGHIWLAFLICIGVGHILTKFPFWARRALVMSGEAYLSYLAAL 267

QY 70 -----QCFSRYPDHMKQHDFFKSAPE-GYVOERTIFFKDD-----GN 106
DB 268 ALMGFIANCFVWFNFNTAYPSEFPFGTGPASQAQATFLVDRDLGANGVSAQPTGLGK 327

QY 107 YKTRA---EVKFEGLTLVNRIELKIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIK 163
DB 328 YLMRSPSGEIIFGGTM-----RFDWTRAPMPLRGANGLD 364

QY 164 VNFKIRHNIEDGSVQLADHYQONTPIGD---GPVLLPDN---HYLSTQSALSKDPNEKRD 217
DB 365 LT-KIKYDIQPMOERRAAEYMTAPLGSLNSVGVATEINSVNVSPRSWLS-----TS 417

QY 218 HMVLGGFV-----TAAGITLGM 235
DB 418 HFVLGFFLFIHGLWHAGRARASGGFEKGLD 448

RESULT 4
H64102
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)
N:Alternate names: leucyl-tRNA synthetase
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
C:Accession: H64102
R; Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: H64102
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-861 <TRIG>
A:Cross-references: GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC22581.1; PID:g1573943; T.
C:Genetics:
A:Gene: leuS
C:Superfamily: leucine-tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.2%; Score 91.5; DB 2; Length 861;
Best Local Similarity 24.1%; Pred. No. 14;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVPWPTLVTLLSYGQCFSRYPDHMKQDFFKSAPEGYVQERTIFFKD----- 103
DB 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH--DQRDF-----EFAQKYSLPKQVIAPLA 364
QY 104 DGNKTRAEVKFEGDGLVNRIELKIDFKEDGNILGHKLEBYNNYNNHNVIMADK-OKNGI 162
DB 365 DEIDLTKQAFVEHGLKLVNSDEFGRNF--DGAENG-----IADKLEKLG 408
QY 163 ---KVNFKIRH-----NIEDGSVOLADHYQONTPIGDPVLLPDNHYL- 202
DB 409 GKQVNYRLRDWGSQRQYWGAPIMPLTLENGDVPA-----PMEDLPILPEDVWMD 461
QY 203 STQSALSQDPN 213
DB 462 GVKSPINADPN 472

RESULT 5
C97354
hypothetical protein CAC3702 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C97354
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97354
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-425 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81622.1; PID:g15026806; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3702

Query Match 7.1%; Score 90; DB 2; Length 425;
Best Local Similarity 24.7%; Pred. No. 7.3;
Matches 66; Conservative 31; Mismatches 98; Indels 72; Gaps 16;

QY 21 GDVNGHKFSVSGEGDATYCKLTLPKICTTK-----LPVPPTLVTLLSYGQC- 71
DB 109 GFVNGKLLPASGEAIKD--FGKV----LKNTGDKLGLTRQVEMFPG-VGRVSVGPNTIE 161
QY 72 FSRYPDHMK-----QHDFFKSAMPEGYVQERTIFFKDDGNKYTR-----AEVKFEGD-- 118
DB 162 FSEVWVNFNGVGDIKDNFVAKVSEDKVMGNP---SGEGNYSTEWEYNLYLKEKYGEDNV 218

QY 119 -----TLVNRIELKGIDPKE--DGNILGHKLEYN-----NSH--NVYIMA 155
DB 219 YLCTDEKSLINSVSEKLSFRKNGVDYV-KTAVAYGENLDTVSMATNSHQYNPILT 277

QY 156 DKQNGIKVNFKIRHIEDGSVOLADHYQONTPIGDPVLLPDNHYLSTQSALSKD----- 211
DB 278 DGEKFDKNMYNKLKEYGDESVAISKY--BHTLTEDIKTLPEYF--TSKGLVKKDDVIS 334

QY 212 -----PNEKRDMVLGGFVTAAG 229
DB 335 EMKKLTDIAIENTNARIAGKFSKLTG 361

RESULT 6
T06586
DNA-binding protein PD2 - garden pea
C;Species: Pisum sativum (garden pea)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C;Accession: T06586
R;Sato, N.; Kazuno, A.A.; Ohta, N.; Oshima, K.

submitted to the EMBL Data Library, June 1996
A;Description: Identification of a novel family of DNA-binding proteins with two AT-hook
A;Reference number: Z15774
A;Accession: T06586
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-632 <SAT>
A;Cross-references: EMBL:X98740; NID:e9952239; PIDN:CAA67292.1; PID:e275185
A;Experimental source: cv. Alaska

Query Match 7.1%; Score 90; DB 2; Length 632;
Best Local Similarity 23.3%; Pred. No. 12;
Matches 49; Conservative 26; Mismatches 79; Indels 56; Gaps 7;

QY 16 LVLEDGDVNGHKFSVSGEGDATYCKLTLPKICTTKLPVPWPTLVTLLSYGQCFSRY 75
DB 363 IVUQGRDVGSKVDVINKESNEATIPENK-----PTEPKLDVEQELAATM----- 408
QY 76 PDHMKQDFFKSAPEGYV-----QERTIFFKDDGNKYTRAEVKFEGDGLVNRIE 125
DB 409 PSSAKVNVLTLDLIVETFPILRSVARTSSGREGSEELKDSGNSLERDTKKLELQGNKS-E 467
QY 126 LKQIDFKEDGNILGHKLE-----YYNSHNVYIMADKQK-----N 160
DB 468 LKGIEPTDSTLLDEKFENALGNKILKEISNPRHDVESANHSHTNKQVTVSHQKAIETNN 527
QY 161 GIKVNFKIRHNIEDG-----SVQLADHYQ 184
DB 528 QSOVEDVAKNKIQDDSKPSEESLHKADKYR 557

RESULT 7
E82590
leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: E82590
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: E82590
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-887 <SIM>
A;Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN0012
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaka
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF2176
C;Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 887;
Best Local Similarity 22.2%; Pred. No. 21;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPWPTLVTLLSYGQCFSRYPDHMKQDFFKSAPEGYVQERTIFFKDDGNY-- 107
DB 329 TNEQLPV-WVANFVLMAYGTGAVMAVPGHQDQRDEF--ANKYGLPIRQVIALKEPKNQDE 385

QY 108 -----KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHVYI 153
DB 386 STWEPDVRWDYAKTR---EPF---LINSAPFDGLDYQDAFEVLAERFE----- 429
QY 154 MADKQKNG-IKVNFKIRHNIEDGSVQLADHYQONTPI-----GDGPVLLPDN 199
DB 430 ---RQGRQRVNYRLR---DWGVSQRVYWGCPPIVYCTCGAVPVPEDQLPVILPEN 482
QY 200 -HYLSTQSALSADPNNEKR 216
DB 483 VAPSGTGSPIKTPDEWRK 500
RESULT 8
DB3917
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: D83917
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: D83917
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-655 <STO>
A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05859.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH2140
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B
Query Match 6.9%; Score 88.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;
QY 22 DVNGHK---FSVSGEGEGDAT---YKLTLEFI-----CTTGKLPVWP 59
DB 63 NVTIHKDQSVSRDEGRGMPGMHKLKGPTEVILTVLHAGKFGGGVATSGLHGVA 122
QY 60 TLVTTLVSYGQCFSRYPDHMKQHDFFKSAMPEGYVOER-----TIFFKDDG----- 105
DB 123 SVVNAISEWLIVIEIKDGVYEOFPENGKGPSTLEKKGKTRGTGTHFKPDPTVFSTT 182
QY 106 --NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADK----- 157
DB 183 NFNVETLSERLEAAFLKGLKIELVDLRDRTKEVFH-YEDGIKAFVEYLNEDKETLHPV 241
QY 158 -----QKNGIKVNFKIRHNIEDGSVQLADHYQONTPIGDGPVLLPDNHYLSTQSALSADPN 212
DB 242 VFPNGESNGIEIFAFQFN--DGYTENVLVSFVNVRTKDG-----GTHLGAKTAMTRAV 294
QY 213 NE 214
DB 295 NE 296
RESULT 9
S41480
Photosystem II chlorophyll a-binding protein psbC - Chlamydomonas eugametos chloroplast
C:Species: chloroplast Chlamydomonas eugametos
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999
C:Accession: S41480
R;Cote, M.J.J.; Lemieux, C.; Turmel, M.
submitted to the EMBL Data Library, May 1993
A;Description: Identification of self-splicing group I introns in the chloroplast psbC g
of Chlamydomonas moewusii.
A;Reference number: S41480
A;Accession: S41480
A;Molecule type: DNA
A;Residues: 1-461 <COT>

A;Cross-references: EMBL:M90639; NID:g289948; PIDN:AAA84148.1; PID:g289949
C;Genetics:
A;Gene: psbC
A;Genome: Chloroplast
A;Start codon: GTG
A;Introns: 181/3; 294/3
C;Superfamily: photosystem II chlorophyll a-binding protein psbC
C;Keywords: chlorophyll; chloroplast; membrane-associated complex; photosynthesis; photosynthesis; photosynthesis
Query Match 6.9%; Score 88; DB 2; Length 461;
Best Local Similarity 19.5%; Pred. No. 12;
Matches 51; Conservative 37; Mismatches 83; Indels 90; Gaps 12;
QY 37 DATYKGLTLKFICTTGKL---PVPWP-----TLVTTLVSYGQCFSR 74
DB 220 DIIGHIWTGTCIFGCIWHIYTPWPWARRAFVMSGEAYLSYSLGAIATMGFIACFSW 279
QY 75 YPDHMKQHDFFKSAMPE-GYVOERTIFFKDD-----GNKYTRA---EVKF 115
DB 280 FNNATVPSEFYGPTGPEASQAQAFPLVRDQLGANVASAQQGTGLGKYLMSRPTGEIIF 339
QY 116 EGDTLVNRTELKIDFKEDGNILGHKLEYNNSHVYIMADKQKGIKVNFKIRHNIEDG 175
DB 340 GGET-----MRFWDFR-----GPWCPLRGFNGLDLN-KLUNDIQPW 375
QY 176 SVQLADHYQONTPIGD---GPVLLPDN---HYLSTQSALSADPNNEKRDMHVLGGF----- 224
DB 376 QERRAAEYTHAPLSGLNSLVGGVATEINAVNFVSPRWLA-----TSHFVLGFFFFVGH 429
QY 225 -----VTAAGITLGMDEL 237
DB 430 LWHAGRARAAAAGFEKGIDRV 450
RESULT 10
I40794
dihydrolipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum
N;Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase
hydrogenase complex chain E3; S-complex 50K chain
C;Species: Clostridium magnum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: I40794
R;Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuchel, A.
J. Bacteriol. 176, 3614-3630, 1994
A;Title: Biochemical and molecular characterization of the Clostridium magnum acetoin de
A;Reference number: I40789; MUID:94266715; PMID:8206840
A;Accession: I40794
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-578 <KRU>
A;Cross-references: GB:L31844; NID:g472324; PIDN:AAA21748.1; PID:g472330
C;Function:
A;Description: catalyzes the oxidation of dihydrolipoamide to lipoamide using NAD
A;Pathway: acetoin dehydrogenase enzyme system
C;Superfamily: Alkaligenes dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase
C;Keywords: FAD; flavoprotein; lipoamide; NAD; oxidoreductase; redox-active disulfide
F;5-77/Domain: lipoyl/biotin-binding homology <LPB>
F;117-145/Region: beta-alphabeta FAD nucleotide-binding fold
F;119-561/Domain: dihydrolipoamide dehydrogenase homology <DLID>
F;287-315/Region: beta-alphabeta NAD nucleotide-binding fold
F;153-158/Disulfide bonds: redox-active #status predicted
Query Match 6.9%; Score 88; DB 1; Length 578;
Best Local Similarity 22.9%; Pred. No. 16;
Matches 57; Conservative 43; Mismatches 91; Indels 58; Gaps 13;
QY 10 TGVVPIVLVDGNGHKSVSVEGEGDATYKLTUK-----FICTGKLPWPWTLVT 63
DB 255 TGSMPPIPIEIE---GNKLS---GVIDST-GALSLESPESIAIIGGGVIGVEFASIFN 305
QY 64 TLSYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNKYKTRAEVKFEGDTLVNR 123
DB 306 SLGCKSVIIEMLPHILPFPMDREISEI-----AKALIRGININN 346

[illegible]

RESULT 15
AE1055
probable exported protein ytfN [imported] - Salmonella enterica subsp. enterica serovar
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AE1055
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608

Search completed: June 21, 2004, 16:02:00
Job time : 11.3333 secs

; Pettea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A;Reference number: A71600; MUID:99021743; PMID:9804551
 A;Accession: D71614
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-2573 <GAR>
 A;cross-references: GB:AE001396; GB:AE001362; NID:G3845188; PIDN:AAC71881.1; PID:G384519
 A;Experimental source: clone 3D7
 C;Genetics:
 A;Gene: PPE0460C

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-222G

Perfect score: 1276

Sequence: 1 MVSKEELFTGVVPIIVELD.....VLGGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	97.1	238	1	GFP_AEQVI
2	92	7.2	461	1	PSBC_CYPAPA
3	91.5	7.2	861	1	SYL_HAEIN
4	89.5	7.0	879	1	SYL_XYLFA
5	88	6.9	461	1	PSBC_CHLEU
6	87.5	6.9	860	1	SYL_SALTI
7	87.5	6.9	860	1	SYL_SALTY
8	87.5	6.9	222	1	DPOE_YEAST
9	87	6.8	689	1	AC2L_HUMAN
10	86.5	6.8	533	1	CP51_CANGA
11	86.5	6.8	879	1	SYL_XYLFT
12	86.5	6.8	1603	1	VIT1_CABEL
13	86	6.7	357	1	TRMA_CAMJE
14	85.5	6.7	353	1	HIS7_BUCAI
15	85.5	6.7	658	1	ADAS_HUMAN
16	85.5	6.7	788	1	DPOL_HBHE
17	85.5	6.7	886	1	ITH3_MSAU
18	85.5	6.7	1259	1	YTFN_ECOLI
19	85	6.7	504	1	YC03_KLEPN
20	84.5	6.6	613	1	PEPF_MYCPU
21	84.5	6.6	859	1	SYL_SHRON
22	84.5	6.6	888	1	LAGC_DICDI
23	84	6.6	366	1	SET7_HUMAN
24	83.5	6.5	538	1	GRBE_RAT
25	83.5	6.5	589	1	SYD_HAEDU
26	83.5	6.5	1164	1	BAG_STRAG
27	83	6.5	472	1	PSBC_SYNY3
28	83	6.5	874	1	SLAP_BACLI
29	82.5	6.5	501	1	AMPA_WTGBR
30	82.5	6.5	533	1	NIFD_CLOPA
31	82.5	6.5	658	1	ADAS_CAVPO
32	82.5	6.5	795	1	D152_HAEIN
33	82.5	6.5	797	1	D151_HAEIN

34	82	6.4	682	1	PRC_ECOLI	P23865	escherichia
35	82	6.4	728	1	CATB_ASPFU	Q92405	aspergillus
36	82	6.4	752	1	NEC1_RAT	P28840	rattus norv
37	81.5	6.4	793	1	D153_HAEIN	O32629	haemophilus
38	81.5	6.4	941	1	GUN_BACS6	P19424	bacillus sp
39	81	6.3	336	1	YD48_METJA	O58743	methanococ
40	81	6.3	682	1	AC2L_MOUSE	Q99nb1	mus muscucu
41	81	6.3	737	1	OPT1_DROME	P91679	drosophila
42	81	6.3	774	1	AMY2_SCHPO	O42918	schizosacch
43	80.5	6.3	393	1	TRMB_HELPY	O25443	helicobacte
44	80.5	6.3	860	1	SYL_ECO57	Q8xbn8	escherichia
45	80.5	6.3	860	1	SYL_ECOL6	Q8fjy9	escherichia

ALIGNMENTS

```

RESULT 1
GFP_AEQVI          STANDARD;          PRT;   238 AA.
ID   GFP_AEQVI
AC   P42212; Q17104; Q27903;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Green fluorescent protein.
GN   GFP.
OS   Aequorea victoria (Jellyfish).
OC   Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC   Aequoreidae; Aequorea.
OX   NCBI_TaxID=6100;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92175527; PubMed=1347277;
RA   Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA   Cormier M.J.;
RT   "Primary structure of the Aequorea victoria green-fluorescent
RT   protein.";
RL   Gene 111:229-233(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94185810; PubMed=8137953;
RA   Inouye S., Tsuji F.I.;
RT   "Aequorea green fluorescent protein. Expression of the gene and
RT   fluorescence characteristics of the recombinant protein.";
RL   FEBS Lett. 341:277-280(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97299832; PubMed=9154981;
RA   Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT   "Enhanced expression in tobacco of the gene encoding green fluorescent
RT   protein by modification of its codon usage.";
RL   Plant Mol. Biol. 33:989-999(1997).
RN   [4]
RP   CHROMOPHORE.
RX   MEDLINE=93192221; PubMed=8448132;
RA   Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;
RT   "Chemical structure of the hexapeptide chromophore of the Aequorea
RT   green-fluorescent protein.";
RL   Biochemistry 32:1212-1218(1993).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=96355665; PubMed=8703075;
RA   Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,
RA   Remington S.J.;
RT   "Crystal structure of the Aequorea victoria green fluorescent
RT   protein.";
RL   Science 273:1392-1395(1996).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=98294543; PubMed=9631087;
RA   Yang F., Moss L.G., Phillips G.N. Jr.;
RT   "The molecular structure of green fluorescent protein.";
RL   Nat. Biotechnol. 14:1246-1251(1996).

```



```
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSXDPNEKRDMVLEFVTAAGITHGMDELYK 238
RESULT 2
PSBC_CYPAC STANDARD; PRT; 461 AA.
AC P48104;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).
GN PSBC.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RA "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RA "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RA the genetic complexity of a primitive plastid.";
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -!- FUNCTION: The 43 kDa protein (p6) is a component of the core of
CC photosystem II. It is a chlorophyll binding protein.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Cyanelle
CC thylakoid membrane.
CC -!- SIMILARITY: Belongs to the psbB / psbC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U30821; AAA81279.1; -.
CC PIR; T06936; T06936
CC InterPro; IPR005869; Photo44.
CC InterPro; IPR000932; PSI1prot.
CC Pfam; PF00421; PSI1; 1.
CC TIGRFAMs; TIGR01153; psbc; 1.
CC Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Cyanelle;
CC Transmembrane.
CC SEQUENCE 461 AA; 50402 MW; 62664E39E13B9C1 CRC64;
Query Match 7.2%; Score 92; DB 1; Length 461;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 58; Conservative 35; Mismatches 88; Indels 90; Gaps 13;
QY 25 GHKFSVSGEGDATYKGLTKLICTTGK-----PVPWP--TLVTT---LSYGV--- 69
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 GDGWISVDNMEDIIGGHIWLPALICIGGWHILTKPFWARRALVWSGEAVLSYSLAAL 267
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 -----QCFSRYPDHMKQHDFFKSAMPE-GYVQERTIFPKD-----GN 106
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 ALMGFIANCVFVFNNTAYSEFFGPTGPEASQAQFTFLVRDQRLGANVGSAGPTGLGK 327
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 107 YKTRA---EVKFFGDTLVNRNIELKGIDFKDGNILGHKLEYNYNHNHVMADKQKNGIK 163
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 328 YLMRSPSGSIIFGGTGM-----RFWDTRAPWLEPLRGANGLD 364
QY 164 VNFKIRHNIEDSGVOLADHYQONTPIGD-----GPVLLPDN---HYLSQTSALSXDPNEKRD 217
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 LT-KIKYDIQWPQERRAAEYMTAPLGSLSNVSQVATEINSVYVSPRWLS-----TS 417
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 HMVLGGFV-----TAAGITLGM 235
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 HFVLGFFLFIQHLWHAGRARAAASGGFEGKGLD 448
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 3
SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInnes K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32774; AAC22581.1; -.
CC PIR; H64102; H64102.
CC TIGR; HI0921; -.
CC HAMAP; MF_00049; -.
CC InterPro; IPR002302; Leu-tRNA-synthet.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS-like edit.
CC Pfam; PF00133; tRNA-synt_1_1.
CC PRINTS; PRO0985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 42 52 "HIGH" REGION.
CC SITE 619 623 "KMSKS" REGION.
CC BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 861 AA; 97750 MW; EB93304F6B4C9FB7 CRC64;
```

Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 7;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVWPPTLTVTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKRD-----EFAQKYSIPKQVIAPLA 364
DB 314 TGDKLPV-VWVNFVLMHYGTGAVMVPAAH-DQRDF-----EFAQKYSIPKQVIAPLA 364

QY 104 DGNKTRAEVKFGDTLVNRIELKGDIFKEDGNILGHKLEYNVNSHNVMADK-QKNGI 162
DB 365 DEBIDLTKQAFVHGKLVNSDEPDGKNF--DCAFNG-----IADKLEKLG 408

QY 163 ---KNFKIRH-----NIEDGSVQLADHYQQNTPIGDGPVLLPDNHYL- 202
DB 409 GKQVNRURDWMGVSQRVYWGAPIMPLTLENGDVVFA-----PWEDLPILPEDVWMD 461

QY 203 STQSALSKDPN 213
DB 462 GVKSPINADPN 472

RESULT 4
SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9PEG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE LEUCYL-tRNA synthetase (SC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUC CR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barro M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Coutinho L.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon L.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.P., Truffi D., Teai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RL "The genome sequence of the plant pathogen Xylella fastidiosa.";
Nature 406:151-159(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004031; AAF84975.1; ALT_INIT.
DR HAMAP; MF 00049; -; 1.
DR InterPro; IPR002302; Leu-trna-syntla.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; Valrs_1fers_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 9796 MW; 9FDCB992092919E CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 10;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVWPPTLTVTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDGNY-- 107
DB 321 TNEQLPV-VWVNFVLMHYGTGAVMVPAGHDQRDEF--ANKYGLPIRQVIALKEPKNQDE 377

QY 108 -----KTRAEVKFGDTLVNRIELKGDIFKEDGNILGHKLEYNVNSHNVI 153
DB 378 STWEPDVWRDWDYADKTR---EFE---LINSAFDGLDYQDAFEVLAERFE----- 421

QY 154 MADKQKNG-IKYNFKIRHNIEDGSVQLADHYQQNTPI-----GGGPVLLPDN 199
DB 422 ---RQGRQRRVNYRLR---DWGVSQRVYWGCPVIVYCTCGAVPVPEDQLPVILPEN 474

QY 200 -HYLSQTQSALSKDPNEKR 216
DB 475 VAFSGTGSPIKTDPEWRK 492

RESULT 5
PSBC_CHLEU STANDARD; PRT; 461 AA.
AC Q08684;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).
GN PSBC.
OS Chlamydomonas eugametos.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3053;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=94077731; PubMed=7504814;
RX Turmel M., Mercier J.P., Cote M.J.J.;
RT "Group I introns interrupt the chloroplast psbA and psbC and the
RT mitochondrial rnl gene in Chlamydomonas.";
RL Nucleic Acids Res. 21:5242-5250(1993).
CC -1- FUNCTION: The 43 kDa protein (p6) is a component of the core of
CC photosystem II. It is a chlorophyll binding protein.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC thylakoid membrane.
CC -1- SIMILARITY: Belongs to the psbB / psbC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation


```
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -I- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE008725; AAL19599.1; -.
CC HAMAP; MF 00049; -.
CC InterPro; IPR002302; Leu-tRNA-synt1a.
CC InterPro; IPR002300; tRNA-synt1a.
CC InterPro; IPR001412; tRNA-synt1.
CC InterPro; IPR009008; ValRS IleRS_edit.
CC Pfam; PF00133; tRNA-synt1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 42 52 "HIGH" REGION.
CC BINDING 619 623 "KMSKS" REGION.
CC BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;
CC
CC Query Match 6.9%; Score 87.5; DB 1; Length 860;
CC Best Local Similarity 23.3%; Pred. No. 15;
CC Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;
CC
QY 50 TTGKLPPWPTLVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFFDGNGYKT 109
DB 314 TGEEIPV-WAANFVLMVEYGTGAVNAVEGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 310 RAEVKFGEDTLVNRLELKGIDFKEDGILGHKEVYNNSHVYIMADKQNGIKVNFKIR 169
DB 371 SEQALTEKGVLFNSGEFDGLAFEAFAFNAIDKL-----AEKGVGERKVNTRLR 418
QY 170 H-----NIDGVSQVLADHYQOQNTPIGDGPVLLPDNHYL-STQSALSKDP 212
DB 419 DWGVSQRQRYGAPIMVYLEDGTV-----LPTPEDQLPVLPEDDVMDGITSPIKADP 471
CC
RESULT 8
ID -DPOE YEAST STANDARD; PRT; 2222 AA.
AC P21951.
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
DE polymerase II subunit A).
GN POL2 OR DIN2 OR YNL262W OR N0825.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.
RA MEDLINE=90381771; PubMed=2169349;
RA Morrison A., Araki H., Clark A.B., Hamatake R.K., Sugino A.;
RT "A third essential DNA polymerase in S. cerevisiae."
RL Cell 62:1143-1151(1990).
RN
RP SEQUENCE OF 1-2221 FROM N.A.
RC STRAIN=S288c / FV1679;
RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT genes."
RT Yeast 12:505-514 (1996).
RL
RN [3]
RP TEMPERATURE SENSITIVE MUTANTS.
RX MEDLINE=92164663; PubMed=1537345;
RA Araki H., Ropp P.A., Johnson A.L., Johnston L.H., Morrison A.,
RA Sugino A.;
RT "DNA polymerase II, the probable homolog of mammalian DNA polymerase
RT epsilon, replicates chromosomal DNA in the yeast Saccharomyces
RT cerevisiae."
RL EMBO J. 11:733-740(1992).
CC -I- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
CC REPLICATION.
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N)
CC -I- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30
CC kDa, AND 29 kDa).
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNITS B AND C.
CC -I- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
CC alpha, beta, gamma, delta, and epsilon which are responsible for
CC different reactions of DNA synthesis.
CC -I- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M60416; AAA88711.1; -.
CC EMBL; X92494; CAA63235.1; -.
CC EMBL; Z71538; CAA96169.1; -.
CC PIR; A36028; A36028.
CC GeneOnline; 143268; -.
CC SGD; S0005206; POL2.
CC GO; GO:0000731; P.DNA repair synthesis; IMP.
CC InterPro; IPR006172; DNA_pol_B.
CC InterPro; IPR006134; DNA_pol_B_dom.
CC InterPro; IPR006133; DNA_pol_B_exo.
CC Pfam; PF00136; DNA_pol_B; 1.
CC Pfam; PF03104; DNA_pol_B_exo; 1.
CC SMART; SM00486; POLBc_1.
CC PROSITE; PS00116; DNA_POLYMERASE_B; FALSE NEG
CC Transferase; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Zinc-finger; Nuclear protein.
CC ZN_FING 2108 2181 POTENTIAL.
CC VARIANT 644 644 M -> I (IN POL2-9 TS MUTANT).
CC VARIANT 710 710 P -> S (IN POL2-18 TS MUTANT).
CC SEQUENCE 2222 AA; 255669 MW; CBDDDE2AB147D65B CRC64;
CC
Query Match 6.9%; Score 87.5; DB 1; Length 2222;
CC Best Local Similarity 28.2%; Pred. No. 46;
CC Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;
CC
QY 54 LPVPWP-TLVTTLSYGVCFSRYPDH-----KQHDFFKSAMPEGVV---QERTI 99
DB 883 LPKSPFETVFTLVENGKLYLSTPCSLNRYVHQKFTNHQYQELKDPVNIYTHSENTI 942
QY 100 FFKDDGNGYKTR--AEVKFEGDTLVNR-----TELKIDFKEDGNLGHKEVYNN 147
DB 943 FFEVDGPGYKAMILPSSKEEGIKRVAVFNEVDSGLAEKGFELKRRGEL---OLIKNFQ 999
```

QY 148 S--HNVYIMAD 156
 Db 1000 SDIFKVFLED 1010

RESULT 9
 AC2L HUMAN STANDARD; PRT; 689 AA.
 ID QSNUB1: Q81V99; Q8N234; Q96J11; Q96JX6; Q9NUB2;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Acetyl-coenzyme A synthetase 2-like, mitochondrial precursor
 DE (EC 6.2.1.1) (Acetyl-CoA ligase 2) (Acetyl-CoA synthetase 2).
 GN ACAS2L OR KIAA1846.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN 1;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leheslahti M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.L.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RL "The DNA sequence and comparative analysis of human chromosome 20.";
 RN Nature 414:665-871(2001).
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Brain, and Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vailion D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fiegler J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [3]
 SEQUENCE OF 169-689 FROM N.A. (ISOFORM 1).
 TISSUE=Placenta, and Tongue;
 RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Irie R., Sato H.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
 RA Kikuchi H., Murakawa K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
 RA Kawakami B., Ota T., Hayashi K., Sugiyama T., Otsuki T., Ishibashi T.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S.,
 RA Kawai Y., Wakamatsu A., Kanehori K., Suzuki K., Sugano S.,
 RA Nagahari K., Masuho Y., Nagai K., Isogai T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 SEQUENCE OF 336-689 FROM N.A. (ISOFORM 1).
 TISSUE=Brain;
 RX MEDLINE=21245130; PubMed=11347906;
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XX.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 8:95-95(2001).
 CC -!- FUNCTION: Converts acetate to acetyl-CoA so that it can be used
 for oxidation through the tricarboxylic cycle to produce ATP and
 CO(2) (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
 acetyl-CoA.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9NUB1-1; Sequences=Displayed;
 CC Name=2;
 CC IsoId=Q9NUB1-2; Sequences=VSP_007249;
 CC Notes=No experimental confirmation available;
 CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
 family.
 CC -!- CAUTION: Ref.1 (CAB81894) sequence differs from that shown due to
 erroneous gene model prediction.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
 frameshift in position 250 and numerous sequencing errors.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL035661; CAB75500.1;
 CC EMBL; AL080312; CAB81884.1; ALT_SEQ.
 CC EMBL; BC039261; AAH39261.1;
 CC EMBL; BC044588; AAH44588.1;
 CC EMBL; AK027817; BAB55390.1; ALT_INIT.
 CC EMBL; AK092295; BAC03853.1; ALT_SEQ.
 CC EMBL; AB058749; BAB47475.1;
 CC Genew; HGNC:16091; ACAS2L.
 CC InterPro: IPR000873; AMP-bind.
 CC Pfam; PF00501; AMP-binding; 1.
 CC PROSITE; P800455; AMP_BINDING; 1.
 KW Ligase; Mitochondrion; Transit
 FT TRANSIT 1 36 MITOCHONDRION (POTENTIAL).
 FT CHAIN 37 689 ACETYL-COENZYME A SYNTHETASE 2-LIKE.
 FT DOMAIN 45 53 POLY-ALA
 FT VARSPPLIC 446 447 Missing (in isoform 2).
 FT /FTIG=VSP_007249.
 FT CONFLICT 277 277 V -> M (IN REF. 2; AAH39261).
 FT CONFLICT 488 488 V -> M (IN REF. 2; AAH44588).
 FT SEQUENCE 689 AA; 74856 MW; 66B84E39302AD08B CRC64;


```
Query Match          6.8%; Score 87; DB 1; Length 689;
Best Local Similarity 24.1%; Pred. No. 13;
Matches 33; Conservative 16; Mismatches 36; Indels 52; Gaps 7;

QY 9 FTGWPIVLVDGVDVNGHKFVSVEGEGDATYKGLTKLFCITCTGKLPVMPVPTLVTTLSYG 68
DB 473 FFGIVPVLMDKGSV-----VEGNSVSGALCIS-----QAWPGMARTI--- 510
QY 69 VQCFSPYDPMKQHDFFKSPMPGPGYVQERTIPFKDDGNKYTRA---EVKFEGLTVNRIE 125
DB 511 -----YGDHQREVDAYFRAYP-GY-----YFTGDGAYRTEGGYQITGRMDVDVI----- 553
QY 126 LKGIKDFKEDGNILGHKL 142
DB 554 -----NISGRL 560

RESULT 10
CP51_CANGA
ID_CP51_CANGA STANDARD; PRT; 533 AA.
AC PS0855; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIA1) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility."
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgerer-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIA1) gene fragment."
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -I- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (by
CC similarity).
CC -I- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L40389; AAB02329.1; -
DR EMBL; S75389; AAB32679.1; -
DR
```

```
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match          6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.4%; Pred. No. 10;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;

QY 25 GHKFSVS---GEGEGDATYKGLTKLFCITCTGKLPVMPVPTLVTTLSYGVCFSRYPDH--M 79
DB 109 GHEFIFNAKLADVSAEAYSHL-----TTPVFGKGVYDCPNRLM 149
QY 80 KQHDFKFSAM-PEGYV-----QERTIFFKDDGNKYTRAEVKFEGLTVNRIELKIGDF 131
DB 150 EQKFKVKGALTKEAFVRYVPLIAEEIYKFRSKNFKNENNSGIVDMVMSQPEM--TIF 207
QY 132 KEDGNILGHKLEYNHYNHVMADKQKNGIKVNFKIRHNIEDSGVOLADHYQONTPIGD 191
DB 208 TASRSLLGKEMRKLDTPAYLYSLDKGFTPINF-VFPNLPLEHYRKRDHAQAIS--- 263
QY 192 GPVLLPDNHYLSQTQSALSKDPEKRD 217
DB 264 -----GTYNLSLIKREKND 278

RESULT 11
SYL_XYLFT
ID_SYL_XYLFT STANDARD; PRT; 879 AA.
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR PD1230.
OS Xylella fastidiosa (strain Temecual / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B., Moon D.H.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., da Silva F.R.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Teai S.M.,
RA Carrier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.B., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gaglioti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.B., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa";
RL J. Bacteriol. 185:1018-1026(2003).
CC -I- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(leu).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```


use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; A8013557; AAO29080.1; ALT_INIT.
DR HAMAP; MF_00049; -. 1.
DR InterPro; IPR002302; Leu-TRNAsynt1a.
DR InterPro; IPR002300; trNA-synt 1a.
DR InterPro; IPR001412; trNA-synt 1.
DR InterPro; IPR009008; ValRS 1leRS_edit.
DR Pfam; PF00133; trNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99823 MW; 4C2BE01B8FDC497E CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 18; Mismatches 69; Indels 57; Gaps 10;
Matches 44; Conservative 28;
SQ SEQUENCE 879 AA; 99823 MW; 4C2BE01B8FDC497E CRC64;
QY 50 TTGKLPVPWPTLVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIF--FKDDG-- 107
DB 321 TNEQLPV-WVANFLVAYGTGAVMVGPHDQDEP--ANKYGLPTROVIALKEPKQODE 377
QY 108 -----KTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNYNHNVYI 153
DB 378 STWEPDVRWDYADKTR---EFE---LNSAEPLDGLDYQAFVLAERFE----- 421
QY 154 MADKQKNG-IKYNFKIRHNIEDGSVOLADHYQONTPI-----GDGPVLLPDN 199
DB 422 ---ROGRQRRVYRLLR-----DMGVSQRQYWGCPPIVYCTGAVVPENQLPVILPEN 474
QY 200 -HYLSQTQSALSKDPNEKR 216
DB 475 VAFSGTGSPIKTDPEWRK 492

RESULT 12
ID VIT4 CAEBL STANDARD; PRT; 1603 AA.
AC P18947; Q98PP3.
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
GN VIT-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T., Spieth J., Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=85269643; PubMed=4022780;
RA Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;
RT "The C. elegans vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes."
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -!- FUNCTION: Precursor of the egg-yolk proteins that are sources of

CC nutrients during embryonic development (Potential).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells building the intestine of adult hermaphroditic individuals; they are contralaterally secreted into the body cavity and subsequently taken up by the gonad.
CC -!- SIMILARITY: Contains 1 VWFD domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AC024137; AAK09074.1; -.
DR EMBL; M11498; AAA28163.1; -.
DR EMBL; X02754; CAA26531.1; -.
DR PIR; A43084; A43084.
DR WormPep; F59D8.2; CE26817.
DR InterPro; IPR001747; Lipid transprt_N.
DR InterPro; IPR001846; VWF_D_
DR Pfam; PF01347; Vitellogenin_N; 1.
DR SMART; SM00638; LPD_N; 1.
DR SMART; SM00216; VND; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1603 VITELLOGENIN 4.
FT DOMAIN 1308 1455 VWFD.
FT CONFLICT 30 30 Y -> V (IN REF. 3).
FT CONFLICT 169 169 L -> V (IN REF. 2).
FT CONFLICT 183 187 EVAYT -> RSLRH (IN REF. 2).
FT CONFLICT 275 275 T -> S (IN REF. 2).
SQ SEQUENCE 1603 AA; 186307 MW; E3031703258C99BB CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 1603;
Best Local Similarity 23.4%; Pred. No. 38;
Matches 52; Conservative 32; Mismatches 69; Indels 69; Gaps 12;
QY 1 MVSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTKLFICTTGKLPVWPPT 60
DB 162 MESDKDSLFFNVHEKTMGDCV---AYTIVQEG-GKTIYTKSVNPKCITR-----PE 211
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEG-VYQERTIF---FKDDG----- 105
DB 212 TAYGLRFGSEC-----KECEKQGFVQPIVYTYTFKNEKLQSEVNSIYT 257
QY 106 -----NYKTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNYNHNVIMAD 156
DB 258 LNVNGQEVVVKSETRAKVTFVEESKINR-EIK-----KVSQPKKEIVYSMENEKLEQ 308
QY 157 KQKNG-----IKYNFKIRHNIEDGSVOLADHYQONTP 188
DB 309 FYKQGDGAKEVNPFAKIEIEQKV-EQLEEIFRQIQEH-EQNTP 348

RESULT 13
ID TRMA_CAMJE STANDARD; PRT; 357 AA.
AC Q9PP92;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (tRNA (M-5-U54)-methyltransferase) (RUMT).
GN TRMA OR CJ0831C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.

```
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -!- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
CC 54 (W-5-U54) in all tRNA (By similarity).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing thymine.
CC -!- SIMILARITY: Belongs to the RNA M5U methyltransferase family. Trna
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL139076; CAB73096.1; -.
DR FIR; G91355; G81355.
DR HAMAP; MF_01011; -.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR001566; TrmA.
DR PROSITE; PS01230; TrmA_1; 1.
DR PROSITE; PS01231; TrmA_2; FALSE_NEG.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
FT DOMAIN 207 213
FT ACT SITE 315 315
FT ACT SITE 315 315
FT SEQUENCE 357 AA; 42276 MW; CEC5328347CEE497 CRC64;
Query Match 6.7%; Score 86; DB 1; Length 357;
Best Local Similarity 24.8%; Pred. No. 6.9;
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;
QY 30 KQDFFKAMPEGYOVERTIFFKDDGNYKTRAEVKF--EGDTIV-----NRTELKG 128
Db 14 EKHSFIKKYKFEYTKDFKLFASKDHYRTRAELSFYHENDILFYAMPDPKSKKKYIIEY 73
QY 129 IDPKED-----GNILGHKLEYNYNHNVMADQKNGIKVNFKIRHNIE 173
Db 74 LDFADEKICAFMPRLLEYLRQDNKLEKL-----FGVEFLTQKQ--LSITLLYHKNIE 125
QY 174 D 174
Db 126 D 126
RESULT 14
HIS7_BUCAI STANDARD; PRT; 353 AA.
AC P57203;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Histidine biosynthesis bifunctional protein hisB [Includes:
DE Histidinol-phosphatase (EC 3.1.3.15); Imidazoleglycerol-phosphate
DE dehydratase (EC 4.2.1.19) (IGPD)].
GN HISB OR BU102.
OS Buchnera aphidicola (subsp. Acyrtosiphon pisum) (Acyrtosiphon pisum
OS symbiotic bacterium)
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: D-erythro-1-(imidazol-4-yl)glycerol 3-
CC phosphate = 3-(imidazol-4-yl)-2-oxopropyl phosphate + H(2)O.
CC -!- CATALYTIC ACTIVITY: L-histidinol-phosphate + H(2)O = L-histidinol
CC + phosphate.
CC -!- PATHWAY: Histidine biosynthesis; sixth step.
CC -!- PATHWAY: Histidine biosynthesis; eighth step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: In the N-terminal section; belongs to the histidinol-
CC phosphatase family.
CC -!- SIMILARITY: In the C-terminal section; belongs to the
CC imidazoleglycerol-phosphate dehydratase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001118; BAB12821.1; -.
DR HAMAP; MF_01022; -.
DR InterPro; IPR006549; HAD-SF-IIIa.
DR InterPro; IPR005954; HisB_N.
DR InterPro; IPR006543; Histidinol-phos.
DR InterPro; IPR00807; IGPD.
DR Pfam; PF00475; IGPD; 1.
DR ProDom; PD002282; IGPD; 1.
DR TIGRFAMs; TIGR01662; HAD-SF-IIIa; 1.
DR TIGRFAMs; TIGR01261; hisB_Nterm; 1.
DR TIGRFAMs; TIGR01656; Histidinol-phos; 1.
DR PROSITE; PS00954; IGP_DEHYDRATASE_1; 1.
DR PROSITE; PS00955; IGP_DEHYDRATASE_2; 1.
KW Histidine biosynthesis; Multifunctional enzyme; Lyase; Hydrolase;
KW Complete proteome.
FT DOMAIN 1 164
FT DOMAIN 165 353
FT SEQUENCE 353 AA; 40835 MW; 903BE7E53AD8DAA0 CRC64;
Query Match 6.7%; Score 85.5; DB 1; Length 353;
Best Local Similarity 24.3%; Pred. No. 7.5; 75; Indels 61; Gaps 11;
Matches 53; Conservative 29; Mismatches 29;
QY 64 TLSYGVQCFSRYPDHMKQHDFFKSAMPEGYOVERTIFF--FKDDGNYKTRAEVKF----- 115
Db 61 TESFPLODFS--TAHLFMLSVPFS---EGVIFDDILICHPFLDDDCVCRPKIKMIEPWL 115
QY 116 -----EGDT---LVNRIELKIDFKEDGNILGHKLEYNYNHNVMADQKKN 160
Db 116 DKIDLKKSIVIGDRDQDMQSLNNLKIKYKEDICNWLHITKY-IKENRYAEIIRRTK 174
QY 161 GKYNPFKIRHNIE-----DGSVQLADHYQOQTFIGDGPVLLPDNHYLSTQSALSKDPNEK 215
Db 175 ETKVSIKWLDLEETS KIDTGVKFFDHMLEQSVHSGICM-----NISVQGDLDID---- 225
QY 216 RDH-----MVLGGFVTAA-----GITLGMDE 236
Db 226 -DHTIEDTGIVLGEALLQALGKNGKLSRFGVLPNDE 262
RESULT 15
ADAS_HUMAN STANDARD; PRT; 658 AA.
ID ADAS_HUMAN
AC O00116;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds
(without alignments)
2458.984 Million cell updates/sec

Title: US-09-887-784-222G

Perfect score: 1276

Sequence: 1 MVSKGELFTGVVPIVLVD.....VLGGFVTAAGITLGMDELK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL.25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_virus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1235	96.8	238	2	Q8GHE2
2	1232	96.6	238	5	Q93125
3	1230	96.4	238	2	Q8GHE4
4	1229	96.3	238	2	Q8GHE3
5	1197	93.8	238	5	Q17105
6	1182	92.6	238	5	Q17106
7	1077	84.4	238	5	Q8WTC6
8	1073	84.1	238	5	Q8WTP95
9	1069	83.1	238	5	Q8WTC4
10	1067	83.6	238	5	Q8WTD0
11	1066	83.5	238	5	Q8WTC8
12	1066	83.5	238	5	Q8WTC9
13	1064	83.4	238	5	Q8WTC7
14	1062	83.2	238	5	Q8WTC5
15	252.5	19.8	225	5	Q95UA7
16	252.5	19.8	225	5	Q720W5

17	247	19.4	225	5	Q963F5	Q963F5	montastraea
18	244.5	19.2	236	5	Q8T6U0	Q8T6U0	dendronepht
19	242.5	19.0	225	5	Q720W9	Q720W9	montastraea
20	240	18.8	225	5	Q816J8	Q816J8	trachyphyl
21	238.5	18.7	266	5	Q9U6Y3	Q9U6Y3	clavularia
22	233	18.3	225	5	Q720W4	Q720W4	montastraea
23	232	18.2	224	5	Q8MU48	Q8MU48	montastraea
24	232	18.2	225	5	Q8T5F1	Q8T5F1	montastraea
25	214	16.8	227	5	Q720W6	Q720W6	montastraea
26	214	16.8	234	5	Q720W7	Q720W7	montastraea
27	212.5	16.7	259	5	Q8MMA2	Q8MMA2	agaricia fr
28	210	16.5	239	5	Q8MMA1	Q8MMA1	agarcia ag
29	209	16.4	234	5	Q8T5F2	Q8T5F2	montastraea
30	209	16.4	234	5	Q8MU47	Q8MU47	montastraea
31	208.5	16.3	229	5	Q9U6Y6	Q9U6Y6	anemonia ma
32	206.5	16.2	232	5	Q9GZ28	Q9GZ28	anemonia su
33	206	16.1	227	5	Q9GZP9	Q9GZP9	montastraea
34	206	16.1	227	5	Q720W8	Q720W8	montastraea
35	205.5	16.1	232	5	Q9GP15	Q9GP15	anemonia su
36	204.5	16.0	214	5	Q86LV7	Q86LV7	meandrina m
37	204	16.0	221	5	Q95P04	Q95P04	goniopora t
38	203.5	15.9	214	5	Q86LV8	Q86LV8	meandrina m
39	203.5	15.9	238	5	Q9BLV9	Q9BLV9	renilla mue
40	202	15.8	227	5	Q95V70	Q95V70	montastraea
41	201.5	15.8	225	5	Q9U6Y8	Q9U6Y8	discosoma s
42	198.5	15.6	222	5	Q72168	Q72168	cerianthus
43	198.5	15.6	225	5	Q8T6T9	Q8T6T9	radianthus
44	198	15.5	235	5	Q8T5F0	Q8T5F0	scolymia cu
45	197.5	15.5	232	5	Q9U6Y7	Q9U6Y7	discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	PRELIMINARY;	PRT;	238 AA.
ID	Q8GHE2		
AC	Q8GHE2;		
DC	01-MAR-2003 (TREMELrel. 23, Created)		
DT	01-MAR-2003 (TREMELrel. 23, Last sequence update)		
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)		
DE	Green fluorescence protein.		
GN	2289GFP.		
OS	Azotobacter vinelandii.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
OC	Pseudomonadaceae; Azotobacter.		
OX	NCBI_TaxID=354;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-DSM2289;		
RA	Koranyi P., Berenyi M., Burg K.;		
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria		
RT	Azomonas and Azotobacter."		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF324408; AAN86140.1; -		
DR	GO; GO:0006091; P.energy pathways; IEA.		
DR	InterPro; IPR009017; GFP_like.		
DR	InterPro; IPR000786; Green_fl_protein.		
DR	Pfam; PF01353; GFP; 1.		
DR	PRINTS; PR01229; GFPLORESCENT.		
DR	ProDom; PD013756; Green_fl_protein; 1.		
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;		

Query Match 96.8%; Score 1235; DB 2; Length 238;

Best Local Similarity 97.5%; Pred. No. 1.7e-94;

Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy	2	VSKGELFTGVVPIVLVDVNGHKFSVSGEGDATYKLTLEICTTGKLPVWPPTL	61
Db	1	MSKGELFTGVVPIVLVDVNGHKFSVSGEGDATYKLTLEICTTGKLPVWPPTL	60
Qy	62	VTTLSVGOCFRYPDHMKOHDFFKSAMPEGVVQERTIFFKDDGNKYKRAEVKFGDTLV	121

```
Db      61 VTTFSYGVCFSRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY      122 NRIELKGIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
Db      121 NRIELKGIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180
QY      182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
Db      181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 2
Q93125 ID Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]_TaxID=6100;
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-Optimized mutants of the green fluorescent protein (GFP).";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000917; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DERA60 CRC64;

Query Match 96.6%; Score 1232; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 3e-94;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGNVGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPILVELDGNVGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db 61 VTTFGYVQCFSRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 3
Q93125 ID Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]_TaxID=354;
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF324406; AAN86138.1; -.
GO; GO:0006091; P:energy pathways; IEA.
InterPro; IPR000917; GFP_like.
InterPro; IPR000786; Green_fl_protein.
Pfam; PF01353; GFP; 1.
PRINTS; PR01229; GFPLORESCENT.
ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.3%; Score 1229; DB 2; Length 238;
```

```
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]_TaxID=116849;
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF324405; AAN86137.1; -.
GO; GO:0006091; P:energy pathways; IEA.
InterPro; IPR000917; GFP_like.
InterPro; IPR000786; Green_fl_protein.
Pfam; PF01353; GFP; 1.
PRINTS; PR01229; GFPLORESCENT.
ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.4%; Score 1230; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 4.4e-94;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGNVGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPILVELDGNVGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db 61 VTTFSYGVCFSRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 4
Q93125 ID Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]_TaxID=354;
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF324406; AAN86138.1; -.
GO; GO:0006091; P:energy pathways; IEA.
InterPro; IPR000917; GFP_like.
InterPro; IPR000786; Green_fl_protein.
Pfam; PF01353; GFP; 1.
PRINTS; PR01229; GFPLORESCENT.
ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.3%; Score 1229; DB 2; Length 238;
```

```
Best Local Similarity 97.1%; Pred. No. 5.4e-94;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTKLFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTKLFICTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120

QY 122 NRTELKGIQDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRTELKGIQDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 5
Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748B44 CRC64;

Query Match 93.8%; Score 1197; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 2.4e-91;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTKLFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGKFSVSQGEKGDATYKGLTKLFICTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120

QY 122 NRTELKGIQDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRTELKGIQDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 6
Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748B44 CRC64;

Query Match 92.6%; Score 1182; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 4.2e-90;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTKLFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGKFSVSQGEKGDATYKGLTKLFICTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120

QY 122 NRTELKGIQDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRTELKGIQDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 7
Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
```

```
DR PRINTS; PRO1229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 688FD75E88926903 CRC64;

Query Match 84.4%; Score 1077; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 2.2e-81;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGIVPVLIELDGVNKGHKFSVRGEGDADYKGLKFICTTGKLPVWPPTL 60

QY 62 VTTLSTGVQCFSPYDPHMKOHDFKSAHPGYYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTLSTGICQFARYPEHMKWDFKSAHPGYYQERTIFFQDDGKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGMDFKEDGNILGHKLEYNHNHNYIMADKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQKDPNEKRDMHVLGGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLGDGPVLLPINHYLSTQTAISKORNETRDMHVFLEFFSACGHTGMDLYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qian Y.X., Pang S.Q.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL; AY013824; AKO2062.1; -.
DR EMBL; AY013821; AKO2059.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;

Query Match 84.1%; Score 1073; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 4.6e-81;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGIVPVLIELDGVNKGHKFSVRGEGDADYKGLKFICTTGKLPVWPPTL 60

QY 62 VTTLSTGVQCFSPYDPHMKOHDFKSAHPGYYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTLSTGICQFARYPEHMKWDFKSAHPGYYQERTIFFQDDGKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGMDFKEDGNILGHKLEYNHNHNYIMADKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQKDPNEKRDMHVLGGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLGDGPVLLPINHYLSTQTAISKORNETRDMHVFLEFFSACGHTGMDLYK 238

us-09-887-784-222g.rspt
```


2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYCKLTFLKFICTIGKLPFVFWPTL 61

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds
(without alignments)
1433.395 Million cell updates/sec

Title: US-09-887-784-222I

Perfect score: 1274

Sequence: 1 MYSKGEELFTGVVILVELD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Genesecp1980s:*
- 2: Genesecp1990s:*
- 3: Genesecp2000s:*
- 4: Genesecp2001s:*
- 5: Genesecp2002s:*
- 6: Genesecp2003as:*
- 7: Genesecp2003bs:*
- 8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1272	99.8	239	5	Aae17518 Enhanced
2	1272	99.8	363	6	AbR40352 Human ami
3	1272	99.8	893	4	AgG5781 Amino aci
4	1272	99.8	1132	4	AgG5782 Amino aci
5	1284	99.2	239	5	Aae17517 Enhanced
6	1261	99.0	239	3	Aab22882 Enhanced
7	1261	99.0	239	3	Aay54349 Amino aci
8	1261	99.0	239	3	Aay79584 EGFP sign
9	1261	99.0	239	4	Aab50804 Jellyfish
10	1261	99.0	239	4	Aab85900 A. victor
11	1261	99.0	239	4	Aab31171 Amino aci
12	1261	99.0	239	5	AgG66198 A. victor
13	1261	99.0	239	5	AgG94444 Protease
14	1261	99.0	239	5	Aae14599 Aequorea
15	1261	99.0	239	6	Aae14958 Aequorea
16	1261	99.0	239	6	AgG79829 Green flu
17	1261	99.0	239	6	AbR83616 Green flu
18	1261	99.0	239	6	Ada38074 Aequorea
19	1261	99.0	239	7	Abu63204 Aequorea
20	1261	99.0	239	7	Adc18358 EGFP (enh
21	1261	99.0	239	7	AbW00914 Aequorea
22	1261	99.0	239	7	AdE28570 Enhanced
23	1261	99.0	246	7	Abm79011 Enhanced
24	1261	99.0	248	5	AgG68319 Jellyfish
25	1261	99.0	259	5	Aau99804 Biomembra

ALIGNMENTS

RESULT 1

AAE17518
ID AAE17518 standard; protein; 239 AA.

XX AAE17518;

DT 22-APR-2002 (first entry)

XX Enhanced F64L-E222G jellyfish green fluorescent protein mutant.

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
KW cellular function; genetic reporter; mutant; Stoke's shift; muten.

XX Aequorea victoria.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 65

FT /note= "Wild type Phe substituted with Leu; This
corresponds to position 64 in the wild type protein"

FT Misc-difference 223

FT /note= "Wild type Glu substituted with Gly; This
corresponds to position 222 in the wild type protein"

XX WO200198338-A2.

PD 27-DEC-2001.

XX 18-JUN-2001; 2001WO-EP006848.

XX 19-JUN-2000; 2000DK-00000953.

PR 20-JUN-2000; 2000US-0212681P.

PR 10-MAY-2001; 2001DK-00000739.

PR 10-MAY-2001; 2001US-0290170P.

XX (BIOI-) BIOIMAGE AS.

XX Bjorn SP, Pagliaro L, Thastrup O;

DR WPI; 2002-098224/13.

DR N-PSDB; AAD28163.

XX Novel fluorescent protein in in vitro assay for measuring protein kinase
activity or dephosphorylation activity, or for measuring protein
redistribution, has a green fluorescent protein with F64L and E222G
mutation.

XX Claim 9; Page 37; 41pp; English.

```
XX CC The invention relates to a fluorescent protein derived from green
XX CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
XX CC at F64L and E222G has a bigger compared to other GFP's making it very
XX CC suitable for high throughput screening due to better resolution. The
XX CC fluorescent protein is useful in invitro assays for measuring protein
XX CC kinase activity or dephosphorylation activity, or for measuring protein
XX CC radiations. The fluorescent protein is useful in studying cellular
XX CC functions in living cells; as protein tags in transgenic animals, living
XX CC and fixed cells; organelle tags, secretion marker and genetic integrity.
XX CC The fluorescent protein is also useful as a cell or organelle integrity
XX CC marker, a marker for changes in cell morphology, as transfection marker,
XX CC and as a marker to be used in combination with fluorescence activated
XX CC cell sorting (FACS). The novel proteins can also be used as reporters to
XX CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
XX CC protein is also useful as markers in transcriptional and translational
XX CC fusions for performing transposon vector mutagenesis and as a reporter
XX CC for bacterial detection. Transposons encoding the fluorescent protein are
XX CC useful for screening promoters and for tagging plasmids and chromosomes.
XX CC The fluorescent protein engineered into the genome of a phage is useful
XX CC for designing diagnostic tool. The present sequence is a DNA encoding
XX CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
XX SQ Sequence 239 AA;

Query Match 99.8%; Score 1272; DB 5; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.5e-122;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHVQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
DB 181 DHVQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239

RESULT 2
ABR40352
ID ABR40352 standard; protein; 363 AA.
XX AC ABR40352;
XX DT 08-JUN-2003 (first entry)
XX DE Human amino acid sequence SEQ ID NO: 6.
XX KW Human; heterologous conjugate; intracellular protein.
XX OS Homo sapiens.
XX OS Aequoria victoria.
XX PN WO20001029827-A2.
XX PD 10-APR-2003.
XX FF 01-OCT-2002; 2002WO-DK000651.
XX FR 01-OCT-2001; 2001DK-00001433.
XX FR 11-OCT-2001; 2001US-0328896P.
XX KW (BIOL-) BIOMAGE AS.
XX PA Terry BR, Nielsen SJ;
XX PI
```

```
XX DR WPI; 2003-430211/40.
XX DR N-PSDB; ACC72604.
XX PT Novel cell for identifying modulators of protein interaction, contains a
XX PT first conjugate comprising anchor protein, second conjugate having type B
XX PT interactor protein and third conjugate with detectable group.
XX PS Disclosure; Page 112-113; 118pp; English.
XX CC The invention relates to a novel cell, comprising three heterologous
XX CC conjugates (HC), a first HC (HC1) comprising an anchor protein that
XX CC specifically binds to an internal structure within the cell conjugated to
XX CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
XX CC type B conjugated to a first protein of interest, and a third HC (HC3)
XX CC comprising a second protein of interest conjugated to detectable group.
XX CC The cell is useful for detecting if a compound disrupts or induces the
XX CC interaction between two intracellular proteins. The cell is also useful
XX CC for screening compounds that modulate the interaction between two
XX CC intracellular proteins. The present sequence is used in the
XX CC exemplification of the invention
XX SQ Sequence 363 AA;

Query Match 99.8%; Score 1272; DB 6; Length 363;
Best Local Similarity 99.6%; Pred. No. 2.8e-122;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
DB 1 MVSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHVQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
DB 181 DHVQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239

RESULT 3
AAG65781
ID AAG65781 standard; protein; 893 AA.
XX AC AAG65781;
XX DT 07-JAN-2002 (first entry)
XX DE Amino acid sequence of HSPDE4A1-E222G fusion protein.
XX KW PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
XX KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX KW fusion protein.
XX OS Homo sapiens.
XX OS Aequorea victoria.
XX PN WO200179526-A2.
XX PD 25-OCT-2001.
XX FF 11-APR-2001; 2001WO-DK000264.
XX FR 17-APR-2000; 2000DK-00000651.
XX FR 29-MAY-2000; 2000DK-00000849.
XX KW (BIOI-) BIOIMAGE AS.
XX PA
```

```
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
XX WPI; 2001-611727/70.
DR N-PSDB; AA166852.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 156-160; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
CC protein
XX
XX Sequence 893 AA;
XX
XX Query Match 99.8%; Score 1272; DB 4; Length 893;
XX Best Local Similarity 99.8%; Pred. No. 1.1e-121;
XX Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MYSKGEELTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTLKFICTTGKLPVPWPT 60
Db 655 MYSKGEELTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTLKFICTTGKLPVPWPT 714
XX
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
XX
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
XX
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
Db 835 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 893
XX
RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA.
XX
AC AAG65782;
XX
XX 07-JAN-2002 (first entry)
XX
XX Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW fusion protein.
XX
XX Homo sapiens.
OS Aequorea victoria.
XX
XX WO200179526-A2.
XX
XX
```

```
PD 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-DK000264.
XX
XX 17-APR-2000; 2000DK-00000651.
PR 29-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
XX WPI; 2001-611727/70.
DR N-PSDB; AA166853.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 162-167; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
CC protein
XX
XX Sequence 1132 AA;
XX
XX Query Match 99.8%; Score 1272; DB 4; Length 1132;
XX Best Local Similarity 99.6%; Pred. No. 1.5e-121;
XX Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MYSKGEELTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTLKFICTTGKLPVPWPT 60
Db 894 MYSKGEELTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTLKFICTTGKLPVPWPT 953
XX
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013
XX
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
XX
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
Db 1074 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 1132
XX
RESULT 5
AAE17517
ID AAE17517 standard; protein; 239 AA.
XX
AC AAE17517;
XX
XX 22-APR-2002 (first entry)
XX
XX Enhanced F64L jellyfish green fluorescent protein mutant.
XX
XX
```

Jellyfish; green fluorescent protein; GFP; protein redistribution; cellular function; genetic reporter; mutant; Stoke's shift; mutin.

Aequorea victoria.

Synthetic.

Key Location/Qualifiers

Misc-difference 65 /note= "Wild type Phe substituted with Leu; This corresponds to position 64 in the wild type protein"

WO200198338-A2.

27-DEC-2001.

18-JUN-2001; 2001WO-EP006848.

19-JUN-2000; 2000DK-00000953.

20-JUN-2000; 2000US-0212681P.

10-MAY-2001; 2001DK-00000739.

10-MAY-2001; 2001US-0290170P.

(BIOI-) BIOIMAGE AS.

Bjorn SP, Pagliaro L, Thastrup O;

WPI; 2002-098224/13.

N-PSDB; AAD28162.

Novel fluorescent protein in vitro assay for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution, has a green fluorescent protein with F64L and E222G mutation.

Example 1; Page 35; 41pp; English.

The invention relates to a fluorescent protein derived from green fluorescent protein (GFP) or its analogue. The GFP containing mutations at F64L and E222G has a bigger compared to other GFP's making it very suitable for high throughput screening due to better resolution. The fluorescent protein is useful in vitro assays for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution. The fluorescent protein is useful in studying cellular functions in living cells; as protein tags in transgenic animals, living and fixed cells; organelle tags, secretion marker and genetic reporter. The fluorescent protein is also useful as a cell or organelle integrity marker, a marker for changes in cell morphology, as transfection marker, and as a marker to be used in combination with fluorescence activated cell sorting (FACS). The novel proteins can also be used as reporters to monitor live or dead biomass of organisms, such as fungi. The fluorescent protein is also useful as markers in transcriptional and translational fusions for performing transposon vector mutagenesis and as a reporter for bacterial detection. Transposons encoding the fluorescent protein are useful for screening promoters and for tagging plasmids and chromosomes. The fluorescent protein engineered into the genome of a phage is useful for designing diagnostic tool. The present sequence is enhanced F64L jellyfish green fluorescent protein (GFP) mutant

Sequence 239 AA;

Query Match 99.2%; Score 1264; DB 5; Length 239;

Best Local Similarity 99.2%; Pred. No. 1e-121;

Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 MWSGELFTGVVPIVLVDGVNGHKFSVSGEGDATYKGLTKFTCTTGKLPVWPWT 60

1 MWSGELFTGVVPIVLVDGVNGHKFSVSGEGDATYKGLTKFTCTTGKLPVWPWT 60

61 LVTTLSYGVCFSRYPDHMKHQDFPKSAMPEGYVQERTIFFKDDGNKYTRAEVFGDTL 120

61 LVTTLSYGVCFSRYPDHMKHQDFPKSAMPEGYVQERTIFFKDDGNKYTRAEVFGDTL 120

121 VNRLEKIDFKEDGNILGHKLEVNVSHNVYIMADQKNGIKVNFKIRNIEDGVSQOLA 180

CC biosensors of the invention can be used to investigate a wide range of
 CC cellular activities and to screen compounds which modulate these
 CC activities. Biosensors containing a recognition site for caspase, for
 CC example, may be used for the screening of compounds which modulate
 CC apoptosis, while biosensors containing other protease recognition sites
 CC may be used for the detection of proteolytic toxins (such as anthrax
 CC lethal factor). The method provides improved target validation and
 CC candidate compound optimisation by combining many cell screening formats
 CC with fluorescence-based molecular reagents and computer-based feature
 CC extraction, data analysis and automation, resulting in increased quantity
 CC and speed of data collection and faster evaluation of drug candidates.
 CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used
 CC as components of biosensor fusion proteins of the invention
 XX
 SQ Sequence 239 AA;

Query Match 99.0%; Score 1261; DB 3; Length 239;
 Best Local Similarity 98.7%; Pred. No. 2.1e-121;
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYKLTLFICTTGTGLPVPWPT 60
 DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYKLTLFICTTGTGLPVPWPT 60
 QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

RESULT 7
 AAY54349
 ID AAY54349 standard; protein; 239 AA.
 XX
 AC AAY54349;
 XX
 DT 06-APR-2000 (first entry)
 DE
 XX Amino acid sequence of the mutant green fluorescent protein EGFP.
 XX Fluorescent protein; green fluorescent protein; emission intensity;
 KW fluorescence; pH detection; pH sensor; EGFP.
 OS Synthetic.
 OS Aequorea victoria.

XX Key Location/Qualifiers
 FH Misc-difference 65 /note= "wild type Phe substituted with Leu"
 FT Misc-difference 66 /note= "wild type Ser substituted with Thr"
 FT Misc-difference 232 /note= "wild type His substituted with Leu"
 FT
 XX WO964592-A2.
 PN
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US012850.
 XX
 PR 09-JUN-1998; 98US-00094359.
 PR 13-OCT-1998; 98US-00172063.
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON STATE.
 XX

PI Tsien RY, Llopis J, Wachter RM;
 XX WPI: 2000-116540/10.
 DR N-PSDB; AA245642.
 XX New functional engineered green fluorescent proteins, used for measuring
 PT the pH in biological samples and cells.
 XX
 PS Disclosure; Page 9; 89pp; English.
 XX
 CC The present sequence represents a functional engineered fluorescent
 CC protein based on the Aequorea green fluorescent protein (GFP). The
 CC emission intensity changes as pH varies between 5 and 10 of the present
 CC protein are novel. The functional engineered fluorescent proteins show
 CC reversible changes in fluorescence over physiological pH ranges. They can
 CC be used for determining the pH of samples and cells. The polynucleotides
 CC can also be used to produce transgenic animals. The fluorescent protein
 CC pH sensors can be delivered to cells in the form of polynucleotides
 CC encoding the protein sensor fused to a targeting signal. The targeting
 CC signal directs the expression of the protein sensors to restricted cell
 CC locations. This makes it possible to measure the pH of a precisely
 CC defined cellular region or organelle
 XX
 SQ Sequence 239 AA;

Query Match 99.0%; Score 1261; DB 3; Length 239;
 Best Local Similarity 98.7%; Pred. No. 2.1e-121;
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYKLTLFICTTGTGLPVPWPT 60
 DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYKLTLFICTTGTGLPVPWPT 60
 QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

RESULT 8
 AAY79584
 ID AAY79584 standard; peptide; 239 AA.
 XX
 AC AAY79584;
 XX
 DT 29-AUG-2000 (first entry)
 DE EGFP signal domain.
 XX
 KW Protease; biosensor; EGFP; signal peptide; cell screening; assay;
 KW analysis; drug discovery.
 XX
 OS Unidentified.
 XX
 PN WO200026408-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 29-OCT-1999; 99WO-US025431.
 XX
 PR 30-OCT-1998; 98US-0106308P.
 PR 26-MAY-1999; 99US-0136078P.
 XX
 XX (CELL-) CELLOMICS INC.
 PA
 XX
 PI Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

17-MAY-2000; 2000WO-US013684.
21-MAY-1999; 99US-00316919.
21-MAY-1999; 99US-00316920.
(REGC) UNIV CALIFORNIA.
Tsien RV, Baird GA;
WPI: 2001-032017/04.
N-PSDB; AAC90488.
Novel fluorescent proteins comprising a sensor protein inserted into them, useful for measuring the response of a sensor biological, chemical, electrical or physiological parameter in vivo or in vitro.
Disclosure; Page 24; 94pp; English.
The present sequence is a fluorescent protein used in the construction of a fluorescent protein indicator. The indicator comprises a sensor polypeptide that is responsive to a chemical, biological, electrical or physiological parameter, and a fluorescence protein functional group. The sensor polypeptide is operatively inserted into the fluorescent moiety. The fluorescent indicator is useful for detecting the presence of a response inducing member in a sample. The method involves contacting the sample with the indicator and detecting a change in fluorescence, in which a change is indicative of the effect of the parameter on the sensor polypeptide. The novel fluorescent proteins are advantageous due to their reduced size as compared to the FRET (fluorescence resonance energy transfer)-based sensors

Query Match 99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.1e-121;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

RESULT 10
AAB85900
ID AAB85900 standard; protein; 239 AA.
XX AAB85900;
AC
DT 30-NOV-2001 (first entry)
XX
DE A. victoria green fluorescent protein (GFP) and linker sequence.
XX
KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;
KW green fluorescent protein.
XX
OS Synthetic.
OS Aequorea victoria.
XX
PN WO200168706-A1.
XX

WPI: 2000-365644/31.
N-PSDB; AAA27573.
Recombinant nucleic acid encoding a protease biosensor useful for fluorescence based cell and molecular biochemical assays for drug discovery comprising three operably linked nucleic acid sequences.
Claim 14; Fig 29A; 218pp; English.
The present sequence is that of the EGFP signal domain, which can be included in novel recombinant protease biosensors (PBs) of the invention. The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87) comprising at least 1 detectable polypeptide signal such as the present sequence; a second domain (see AAY79588-622) comprising at least 1 protease recognition site; and a third domain (see AAY79623-37) comprising at least 1 reactant target sequence. A recombinant nucleic acid (see AAA27627-43) encoding the PB, an expression vector, and a genetically engineered host cell are also claimed. A claimed method for identifying compounds that modify protease activity in a cell involves contacting a host cell that possesses the recombinant PB with a test compound, and determining the PB distribution in the host cell, where changes in the distribution of the PB are correlated with modification of protease activity by the test compound. Claimed kits for identifying compounds that modify protease activity in a host cell include the recombinant nucleic acid, or the recombinant PB, or the vector, or the host cell. The PB is useful in high content screens to detect in vivo activation of enzymatic activity, and to identify specific activity based on cleavage of a known recognition motif

Query Match 99.0%; Score 1261; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.1e-121;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

RESULT 9
AAB50804
ID AAB50804 standard; protein; 239 AA.
XX AAB50804;
AC
DT 14-MAR-2001 (first entry)
XX
DE Jellyfish GFP mutant EGFP.
XX
KW Aequorea victoria; jellyfish; fluorescent protein indicator;
KW green fluorescent protein; GFP; linker moiety; sensor;
KW calmodulin-binding domain; mutant; mutin.
XX
OS Aequorea victoria.
XX
PN WO200071565-A2.
XX
PD 30-NOV-2000.
XX

PD	20-SEP-2001.	XX	07-JUN-1999; 99FI-00001296.	XX
XX	14-MAR-2001; 2001WO-US008071.	XX	(LILI/) LILIUS E.	XX
XX	15-MAR-2000; 2000US-0189698P.	PA	(VIRT/) VIRT M.	PA
XX	(MERI) MERCK & CO INC.	XX	Lilius E, Virta M;	XX
XX	Marsh DJ;	PI	WPI; 2001-061737/07.	PI
XX	WPI; 2001-565791/63.	XX	N-PSDB; AAC86954.	XX
DR	N-PSDB; AAH47304.	DR	Assessing growth and death rates of a micro-organism in a desired	DR
XX	Fusion proteins comprising melanin concentrating hormone receptor	PT	environment, by introducing 2 reporter genes encoding luminescent and	PT
PT	peptides and fluorescent proteins, useful for identifying appetite	PT	fluorescent products and detecting luminescent fluorescence.	PT
PT	stimulants.	XX	Disclosure; Page 27; 32pp; English.	XX
XX	Claim 2; Page 14; 71pp; English.	PS		PS
XX	The invention provides melanin concentrating hormone (MCH) receptor	CC	The specification describes a method for assessing the growth rate and	CC
CC	(MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise	CC	death rate of a micro-organism within a predetermined time period in a	CC
CC	MCHR polypeptide regions from different species. The MCHR fusion protein	CC	desired environment. The method comprises introducing at least two	CC
CC	comprise MCHR polypeptide region and a fluorescent polypeptide region	CC	reporter genes encoding luminescent and/or fluorescent products into the	CC
CC	joined directly, or via a linker, to the carboxy side of the MCHR	CC	micro-organisms, incubating the micro-organism within the desired	CC
CC	polypeptide region. The MCHR fusion proteins can be expressed by standard	CC	environment, and detecting luminescence and/or fluorescence after a	CC
CC	recombinant methodology. MCH action promotes feeding (orexigenic) and up	CC	predetermined time period. Use of two different markers within a micro-	CC
CC	regulation of MCH activity stimulates food intake. The present sequence	CC	organism enables the differentiation between growth and death rates. The	CC
CC	represents a A. victoria green fluorescent protein (GFP) and a linker	CC	method is used to assess the growth rate and death rate of a micro-	CC
CC	sequence	CC	organism within a predetermined time period in a desired environment. The	CC
XX	Sequence 239 AA;	CC	present sequence represents a green fluorescent protein (GFP), and is	CC
SQ		CC	encoded by a plasmid which encodes luminescent and fluorescent proteins,	CC
		CC	and is used in the method of the invention	CC
		XX	Sequence 239 AA;	XX
			Query Match 99.0%; Score 1261; DB 4; Length 239;	
			Best Local Similarity 98.7%; Pred. No. 2.1e-121;	
			Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;	
QY	1 MVSKEELFTGVVPILVELDGDVNGHKFVSQEGEGDATYGLTKLFICTTGKLPVPWPT 60	QY	1 MVSKEELFTGVVPILVELDGDVNGHKFVSQEGEGDATYGLTKLFICTTGKLPVPWPT 60	QY
Db	1 MVSKEELFTGVVPILVELDGDVNGHKFVSQEGEGDATYGLTKLFICTTGKLPVPWPT 60	Db	1 MVSKEELFTGVVPILVELDGDVNGHKFVSQEGEGDATYGLTKLFICTTGKLPVPWPT 60	Db
QY	61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	QY	61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	QY
Db	61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	Db	61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	Db
QY	121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180	QY	121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180	QY
Db	121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180	Db	121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180	Db
QY	181 DHYQNTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDHVMVLCFVTAAGITILGMDELYK 239	QY	181 DHYQNTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDHVMVLCFVTAAGITILGMDELYK 239	QY
Db	181 DHYQNTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDHVMVLCFVTAAGITILGMDELYK 239	Db	181 DHYQNTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDHVMVLCFVTAAGITILGMDELYK 239	Db
RESULT 11		RESULT 12		
AAB31171		AAG66198		
ID	AAB31171 standard; protein; 239 AA.	ID	AAG66198 standard; protein; 239 AA.	ID
XX	AAB31171;	XX	AAG66198;	XX
AC		AC		AC
XX	02-APR-2001 (first entry)	XX	17-JUN-2002 (first entry)	XX
DT		DT		DT
XX	Amino acid sequence of a green fluorescent protein (GFP).	DE	A. victoria green fluorescent protein (EGFP).	DE
XX	Growth rate; death rate; reporter gene; luminescent protein;	XX	Cyan-green fluorescent protein; fluorescence; recombinant; GFP;	XX
KW	fluorescent product; luciferase; green fluorescent protein; GFP.	KW	green fluorescent protein; EGFP.	KW
XX	Aequorea victoria.	XX	Aequorea victoria.	XX
OS		OS		OS
XX	WO200075367-A1.	XX	JP2002045189-A.	XX
PN		PN		PN
XX	14-DEC-2000.	XX	12-FEB-2002.	XX
PD		PD		PD
XX	07-JUN-2000; 2000WO-FI000507.	XX		XX

PPF	04-AUG-2000; 2000JP-00237165.
XX	
PR	04-AUG-2000; 2000JP-00237165.
XX	(RIKA) RIKAGAKU KENYUSHO.
PA	
DR	WPI; 2002-299190/34.
DR	N-PSDB; ABL40628.
PT	A gene encoding cyan-green fluorescent protein.
XX	
PS	Example; Page 14; 20pp; Japanese.
CC	The invention relates to a gene encoding proteins having cyan-green fluorescence characteristic and having a function of showing stable CC fluorescence characteristic in acid region. A method for the preparation of a cyan-green fluorescent protein is provided which involves a CC transformant transformed by a recombinant vector comprising the gene, CC where the transformant is cultured and the protein is collected from the CC culture. The present sequence represents the A. victoria green CC fluorescent protein (EGFP)
XX	
SQ	Sequence 239 AA;
	Query Match 99.0%; Score 1261; DB 5; Length 239;
	Best Local Similarity 98.7%; Pred. No. 2.1e-121;
	Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY	1 MVSKEELFTGVVPIVLVDGDNVNGHKFSVSGEGDATYGKLTLCFKICTTGKLVPWPWT 60
Db	1 MVSKEELFTGVVPIVLVDGDNVNGHKFSVSGEGDATYGKLTLCFKICTTGKLVPWPWT 60
QY	61 LVTTLSYGVCFSRYPDHMKQHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
Db	61 LVTTLTGYQCFSRYPDHMKQHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY	121 VNRIELKGIDFDKEDGNILGHKLEYNHNHYIMADKOKNGIKVNFKIRHNIEDGSVQLA 180
Db	121 VNRIELKGIDFDKEDGNILGHKLEYNHNHYIMADKOKNGIKVNFKIRHNIEDGSVQLA 180
QY	181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHWLIGFVTAAGITLGMDELYK 239
Db	181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHWLLEFVTAAGITLGMDELYK 239
RESULT 13	
ABG9444	
ID	ABG9444 standard; protein; 239 AA.
XX	
AC	ABG9444;
XX	
DT	27-NOV-2002 (first entry)
XX	
DE	Protease biosensor signal sequence #6.
XX	
KW	Detection; classification; identification; toxin detection; protease;
KW	App-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
KW	toxic threat agent.
OS	Synthetic.
XX	
PN	US6415959-B1.
XX	
PD	09-JUL-2002.
XX	
PF	25-FEB-2000; 2000US-00513783.
XX	
FR	27-FEB-1997; 97US-00810983.
PR	27-FEB-1998; 98US-00031271.
PR	26-FEB-1999; 99US-0122152P.
PR	08-MAR-1999; 99US-0123399P.
FR	12-JUL-1999; 99US-00352171.
OS	99US-0123399P.

PR	17-SEP-1999; 99US-00398965.
PR	29-OCT-1999; 99US-00430556.
PR	01-DEC-1999; 99US-0168408P.
XX	
PA	(GIUL/) GIULIANO K.
PA	(KAPU/) KAPUR R.
XX	
PI	Giuliano K, Kapur R;
XX	
DR	WPI; 2002-634730/68.
DR	N-PSDB; ABS71491.
XX	
PT	Automated cell-based toxin detection, classification, and/or
PT	identification by treating cells involves use of three classes of
PT	luminescent reporter molecules such as detectors, classifiers or
PT	identifiers.
XX	
PS	Example 10; Fig 29A; 214pp; English.
XX	
CC	The invention describes methods of automated detection, classification
CC	and identification comprising treating cells containing luminescent
CC	reporter molecules (I) in array of locations with a test substance, where
CC	(I) are detectors, classifiers or identifiers imaging cells in each
CC	location to obtain luminescent signals and converting optical information
CC	into digital data to interpret presence of toxins in the test substance
CC	The method are useful for detection of toxins chosen from processes, ADP-
CC	ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
CC	Three classes of cell-based luminescent reporter molecules such as
CC	detectors, classifiers and identifiers are described and serve as
CC	reporters of toxic threat agents. The first two levels of
CC	characterisation ensure a rapid readout of toxin class without
CC	sacrificing the ability to detect many new mutant toxins or dissect
CC	several complex mixtures of known toxins. This is the amino acid sequence
CC	of a protease biosensor related signal sequence used in the cell-based
CC	screening system
XX	
SQ	Sequence 239 AA;
	Query Match 99.0%; Score 1261; DB 5; Length 239;
	Best Local Similarity 98.7%; Pred. No. 2.1e-121;
	Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY	1 MVSKEELFTGVVPIVLVDGDNVNGHKFSVSGEGDATYGKLTLCFKICTTGKLVPWPWT 60
Db	1 MVSKEELFTGVVPIVLVDGDNVNGHKFSVSGEGDATYGKLTLCFKICTTGKLVPWPWT 60
QY	61 LVTTLSYGVCFSRYPDHMKQHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
Db	61 LVTTLTGYQCFSRYPDHMKQHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY	121 VNRIELKGIDFDKEDGNILGHKLEYNHNHYIMADKOKNGIKVNFKIRHNIEDGSVQLA 180
Db	121 VNRIELKGIDFDKEDGNILGHKLEYNHNHYIMADKOKNGIKVNFKIRHNIEDGSVQLA 180
QY	181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHWLIGFVTAAGITLGMDELYK 239
Db	181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHWLLEFVTAAGITLGMDELYK 239
RESULT 14	
AAE14599	
ID	AAE14599 standard; protein; 239 AA.
XX	
AC	AAE14599;
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Aequorea victoria enhanced green fluorescent protein.
XX	
KW	Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; muten.
XX	
OS	Aequorea victoria.
OS	Synthetic.

XX Key Location/Qualifiers
 FT Misc-difference 1. .3
 FT /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"
 FT
 FT Misc-difference 65
 FT /note= "GFP phe64 is replaced by Leu"
 FT Misc-difference 66
 FT /note= "GFP Ser65 is replaced by Thr"
 XX
 XX EP1178109-A1.
 XX
 XX 06-FEB-2002.
 XX
 XX 03-AUG-2001; 2001EP-00306650.
 XX
 XX 04-AUG-2000; 2000JP-00237166.
 PR
 XX (RIKE) RIKEN KK.
 XX
 XX Miyawaki A, Sawano A;
 PI
 DR WPI; 2002-208112/27.
 DR N-PSDB; AAD27910.
 XX
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimer.
 PT
 PT
 PT
 XX
 PS Example 1; Page 13-14; 31pp; English.
 XX
 CC The invention relates to a method for mutagenesis that comprises synthesising a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria
 XX
 SQ Sequence 239 AA;

Query Match 99.0%; Score 1261; DB 5; Length 239;
 Best Local Similarity 98.7%; Pred. No. 2.1e-121;
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
 DB 1 MWSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGDIFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
 DB 121 VNRIELKGDIFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
 QY 181 DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

QY 181 DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

RESULT 15
 AAE34958
 ID AAE34958 standard; protein; 239 AA.
 XX
 AC AAE34958;
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Aequorea victoria enhanced green fluorescent protein (EGFP).
 XX
 XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;
 KW kinase; enhanced green fluorescent protein; EGFP.
 XX
 OS Aequorea victoria.
 XX
 PN WO200295058-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 24-MAY-2002; 2002WO-US016955.
 XX
 PR 24-MAY-2001; 2001US-00865291.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Tsien RY, Ting AY, Zhang J;
 XX
 DR WPI; 2003-148474/14.
 DR N-PSDB; AAD53428.
 XX
 PT Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.
 PT
 XX Disclosure; Col 56-57; 38pp; English.
 PS
 CC The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or in operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention
 XX
 SQ Sequence 239 AA;

Query Match 99.0%; Score 1261; DB 6; Length 239;
 Best Local Similarity 98.7%; Pred. No. 2.1e-121;
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
 DB 1 MWSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGDIFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
 DB 121 VNRIELKGDIFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
 QY 181 DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

us-09-887-784-222i.rag

Sun Jun 27 18:27:31 2004

Db 181 DHYQONTPIGDFVLLPDNHYLSTOSALSKDPNEKRDHNVLLFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:11
Job time : 47.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-222I

Perfect score: 1274

Sequence: 1 MWSKGBELFTGVVPIVLVELD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pdp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pdp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pdp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pdp.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pdp.*
- 6: /cgn2_6/ptodata/2/iaa/backfileesl.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1261	99.0	239	3	US-09-172-063-3
2	1261	99.0	239	4	US-09-513-783A-46
3	1261	99.0	239	4	US-09-316-919-4
4	1261	99.0	239	4	US-09-602-641-3
5	1261	99.0	239	4	US-09-920-922-2
6	1261	99.0	281	3	US-09-062-102-1
7	1261	99.0	281	4	US-09-364-946-1
8	1261	99.0	294	4	US-09-513-783A-2
9	1261	99.0	323	3	US-09-172-063-21
10	1261	99.0	323	4	US-09-602-641-21
11	1261	99.0	364	3	US-09-085-305-6
12	1261	99.0	379	4	US-09-417-197-129
13	1261	99.0	434	4	US-09-800-170-48
14	1261	99.0	442	4	US-09-417-197-127
15	1261	99.0	459	4	US-09-513-783A-170
16	1261	99.0	544	4	US-09-417-197-113
17	1261	99.0	544	4	US-09-417-197-115
18	1261	99.0	604	4	US-09-417-197-59
19	1261	99.0	605	4	US-09-417-197-41
20	1261	99.0	606	4	US-09-417-197-65
21	1261	99.0	607	4	US-09-417-197-47
22	1261	99.0	630	4	US-09-417-197-63
23	1261	99.0	631	4	US-09-417-197-39
24	1261	99.0	633	4	US-09-417-197-45
25	1261	99.0	635	4	US-09-417-197-125
26	1261	99.0	642	2	US-08-818-253-2
27	1261	99.0	642	2	US-08-818-253-6

ALIGNMENTS

RESULT 1									
US-09-172-063-3									
; Sequence 3, Application US/09172063									
; Patent No. 6150176									
; GENERAL INFORMATION:									
; APPLICANT: Tsien, Roger Y.									
; APPLICANT: Miyawaki, Atsushi									
; APPLICANT: Liopis, Juan									
; APPLICANT: Wachter, Rebekka M.									
; APPLICANT: Remington, S. James									
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR									
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE									
; FILE REFERENCE: 07257/071001									
; CURRENT APPLICATION NUMBER: US/09/172,063									
; EARLIER FILING DATE: 1998-10-13									
; EARLIER APPLICATION NUMBER: 09/094,359									
; EARLIER FILING DATE: 1998-06-09									
; NUMBER OF SEQ ID NOS: 36									
; SOFTWARE: FastSeq for Windows Version 4.0									
; SEQ ID NO 3									
; LENGTH: 239									
; TYPE: PRT									
; ORGANISM: Aequorea victoria									
; FEATURE:									
; NAME/KEY: VARIANT									
; LOCATION: (0)...(0)									
; OTHER INFORMATION: EGFP									
US-09-172-063-3									
Query Match 99.0%; Score 1261; DB 3; Length 239;									
Best Local Similarity 98.7%; Pred. No. 1.9e-127;									
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;									
Qy	1	MWSKGBELFTGVVPIVLVELDGVNGHKFSVSGEGDATYVKLTAKFKICTTGKLPVPWPT	60						
Db	1	MWSKGBELFTGVVPIVLVELDGVNGHKFSVSGEGDATYVKLTAKFKICTTGKLPVPWPT	60						
Qy	61	LVTTLSYGVQCSRYRYPDHMKQHDFFKFSAMPEGVYQVQRTIFFKDDGNKYKTRAEVKFEGDTL	120						
Db	61	LVTTLYGVQCSRYRYPDHMKQHDFFKFSAMPEGVYQVQRTIFFKDDGNKYKTRAEVKFEGDTL	120						
Qy	121	VRIELKIDFDKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180						
Db	121	VRIELKIDFDKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA	180						
Qy	181	DHYQONTPTGDGVPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK	239						
Db	181	DHYQONTPTGDGVPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK	239						

Sequence 2, Appli
Sequence 6, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 75, Appli
Sequence 51, Appli
Sequence 139, App
Sequence 176, App
Sequence 141, App
Sequence 143, App
Sequence 77, Appli
Sequence 178, App
Sequence 53, Appli
Sequence 61, Appli
Sequence 43, Appli
Sequence 117, App
Sequence 119, App

RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46
Query Match 99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4
Query Match 99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3
Query Match 99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3
Query Match 99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-920-922-2

Query Match          99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||
Db 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
   |||||
Db 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
   |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
   |||||
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
   |||||

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangliang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          99.0%; Score 1261; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 2.4e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||
Db 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
   |||||
Db 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
   |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
   |||||
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
   |||||
```

```

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangliang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          99.0%; Score 1261; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 2.4e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||
Db 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
   |||||
Db 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120
   |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
   |||||
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
   |||||

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match          99.0%; Score 1261; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 2.6e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||
```

Db 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYKLTLCFKICTTGKLPVWPWT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 61 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLIGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLIGFVTAAGITLGMDELYK 239
RESULT 9
US-09-172-663-21
; Sequence 21, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-172-663-21
Query Match 99.0%; Score 1261; DB 3; Length 323;
Best Local Similarity 98.7%; Pred. No. 3e-127;
Matches 236; Conservative 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYKLTLCFKICTTGKLPVWPWT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYKLTLCFKICTTGKLPVWPWT 144
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 145 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 204
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 264
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLIGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLIGFVTAAGITLGMDELYK 323
RESULT 10
US-09-602-641-21
; Sequence 21, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-602-641-21
Query Match 99.0%; Score 1261; DB 4; Length 323;
Best Local Similarity 98.7%; Pred. No. 3e-127; 1; Indels 0; Gaps 0;
Matches 236; Conservative 2; Mismatches 1;
QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYKLTLCFKICTTGKLPVWPWT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYKLTLCFKICTTGKLPVWPWT 144
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 145 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 204
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 264
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLIGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLIGFVTAAGITLGMDELYK 323
RESULT 11
US-09-085-305-6
; Sequence 6, Application US/09085305
; Patent No. 6191269
; GENERAL INFORMATION:
; APPLICANT: Pollock, Allan
; APPLICANT: Lovett, David H.
; APPLICANT: Turck, Johanna
; TITLE OF INVENTION: Selective Induction of Apoptosis in
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESS: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,305
; FILING DATE: 29-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Francis, Carol L
REGISTRATION NUMBER: 36,513
REFERENCE/DOCKET NUMBER: 6510/102US1
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-085-305-6

Query Match 99.0%; Score 1261; DB 3; Length 364;
Best Local Similarity 98.7%; Pred. No. 3.6e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Db |||||
QY 126 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 185
Db |||||
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db |||||
QY 186 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 245
Db |||||
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db |||||
QY 246 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 305
Db |||||
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
Db |||||
QY 306 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 364
Db |||||

RESULT 12

US-09-417-197-129
Sequence 129, Application US/09417197
Patent No. 6518021

GENERAL INFORMATION:

APPLICANT: Ole THASTRUP, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE OF INVENTION: On A Cellular Response
FILE REFERENCE: 3759-0110P
CURRENT APPLICATION NUMBER: US/09/417,197
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 129
LENGTH: 379
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: actin-binding-domain-EGFP fusion
US-09-417-197-129

Query Match 99.0%; Score 1261; DB 4; Length 379;
Best Local Similarity 98.7%; Pred. No. 3.8e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Db |||||
QY 141 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 200
Db |||||
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db |||||
QY 201 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 260
Db |||||
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db |||||
QY 261 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 320
Db |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
Db |||||
QY 321 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 379
Db |||||

RESULT 13

US-09-800-170-48
Sequence 48, Application US/09800170
Patent No. 6481667

GENERAL INFORMATION:

APPLICANT: Kinsella, Todd
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
FILE REFERENCE: A-68614-1/DJB/RMS/RMK
CURRENT APPLICATION NUMBER: US/09/800,170
CURRENT FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: US 60/187,130
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 90
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48
LENGTH: 434
TYPE: PRT
ORGANISM: Synecocystis PCC6803
US-09-800-170-48

Query Match 99.0%; Score 1261; DB 4; Length 434;
Best Local Similarity 98.7%; Pred. No. 4.7e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Db |||||
QY 196 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 255
Db |||||
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db |||||
QY 256 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 315
Db |||||
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db |||||
QY 316 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 375
Db |||||
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
Db |||||
QY 376 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 434
Db |||||

RESULT 14

US-09-417-197-127
Sequence 127, Application US/09417197
Patent No. 6518021

GENERAL INFORMATION:

APPLICANT: Ole THASTRUP, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE OF INVENTION: On A Cellular Response
FILE REFERENCE: 3759-0110P
CURRENT APPLICATION NUMBER: US/09/417,197
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 127
LENGTH: 442
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EGFP-RhoA fusion
US-09-417-197-127

Query Match 99.0%; Score 1261; DB 4; Length 442;
Best Local Similarity 98.7%; Pred. No. 4.8e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
Db |||||

Db 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTFLKFCITCTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRLELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239

RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliani, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170

Query Match 99.0%; Score 1261; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 5.1e-127; Indels 0; Gaps 0;
Matches 236; Conservative 2; Mismatches 1;
QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTFLKFCITCTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTFLKFCITCTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRLELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:03
Job time : 12.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds
(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-2221
Perfect score: 1274
Sequence: 1 MYSKGBELTGVVPIILVELD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb.p*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb.p*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pdb.p*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pdb.p*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pdb.p*
- 6: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pdb.p*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pdb.p*
- 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pdb.p*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pdb.p*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pdb.p*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pdb.p*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pdb.p*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pdb.p*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pdb.p*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pdb.p*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pdb.p*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pdb.p*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pdb.p*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1272	99.8	239	9	US-09-887-784-4
2	1272	99.8	239	12	US-10-296-953-4
3	1272	99.8	363	14	US-10-270-223-6
4	1272	99.8	893	14	US-10-257-909A-30
5	1272	99.8	1132	14	US-10-257-909A-32
6	1264	99.2	239	9	US-09-887-784-2
7	1264	99.2	239	12	US-10-296-953-2
8	1261	99.0	239	9	US-09-920-922-2
9	1261	99.0	239	9	US-09-999-745-4
10	1261	99.0	239	10	US-09-866-538-4
11	1261	99.0	239	10	US-09-797-496B-2
12	1261	99.0	239	10	US-09-794-308-4
13	1261	99.0	239	10	US-09-865-291-4
14	1261	99.0	239	12	US-10-457-982-3
15	1261	99.0	239	14	US-10-121-258-13

16	1261	99.0	239	14	US-10-221-461-7	Sequence 7, Appli
17	1261	99.0	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1261	99.0	239	14	US-10-177-390-2	Sequence 2, Appli
19	1261	99.0	239	14	US-10-338-411-3	Sequence 3, Appli
20	1261	99.0	239	15	US-10-370-570-4	Sequence 4, Appli
21	1261	99.0	239	15	US-10-389-640-3	Sequence 3, Appli
22	1261	99.0	259	14	US-10-314-861-11	Sequence 11, Appli
23	1261	99.0	281	12	US-09-931-232-1	Sequence 1, Appli
24	1261	99.0	288	14	US-10-314-861-37	Sequence 37, Appli
25	1261	99.0	293	14	US-10-314-861-35	Sequence 35, Appli
26	1261	99.0	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1261	99.0	295	14	US-10-314-861-39	Sequence 39, Appli
28	1261	99.0	299	14	US-10-314-861-33	Sequence 33, Appli
29	1261	99.0	305	14	US-10-314-861-31	Sequence 31, Appli
30	1261	99.0	308	14	US-10-033-717-35	Sequence 35, Appli
31	1261	99.0	311	14	US-10-314-861-29	Sequence 29, Appli
32	1261	99.0	320	14	US-10-338-411-11	Sequence 11, Appli
33	1261	99.0	320	15	US-10-389-640-11	Sequence 7, Appli
34	1261	99.0	323	12	US-10-457-982-21	Sequence 21, Appli
35	1261	99.0	323	14	US-10-338-411-7	Sequence 7, Appli
36	1261	99.0	323	15	US-10-389-640-7	Sequence 13, Appli
37	1261	99.0	323	15	US-10-389-640-13	Sequence 16, Appli
38	1261	99.0	324	14	US-10-338-411-5	Sequence 5, Appli
39	1261	99.0	345	14	US-10-338-411-5	Sequence 9, Appli
40	1261	99.0	345	15	US-10-389-640-5	Sequence 9, Appli
41	1261	99.0	346	14	US-10-338-411-9	Sequence 9, Appli
42	1261	99.0	346	15	US-10-389-640-9	Sequence 33, Appli
43	1261	99.0	359	14	US-10-033-717-33	Sequence 34, Appli
44	1261	99.0	359	14	US-10-033-717-33	
45	1261	99.0	359	14	US-10-033-717-33	

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BUORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequoria Victoria
US-09-887-784-4

Query Match 99.8%; Score 1272; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.2e-124;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MYSKGBELTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLCFICTTGLPVPWPT	60
DB	1	MYSKGBELTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLCFICTTGLPVPWPT	60
QY	61	LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL	120
DB	61	LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL	120
QY	121	VNRIELKGIDFDEKNGILGHKLEYNHNHNYIMADQKNGIKVNFKIRHNIEDGSVOLA	180
DB	121	VNRIELKGIDFDEKNGILGHKLEYNHNHNYIMADQKNGIKVNFKIRHNIEDGSVOLA	180
QY	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK	239
DB	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK	239

```
RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match          99.8%; Score 1272; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.2e-124;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPICCTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPICCTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFRSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFRSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239

RESULT 3
US-10-270-223-6
; Sequence 6, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: Bioimage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM
; TITLE OF INVENTION: INTERACTIONS BY FLUORESCENCE REDISTRIBUTION.
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match          99.8%; Score 1272; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 4e-124;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPICCTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPICCTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFRSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFRSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.8%; Score 1272; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 1.4e-123;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPICCTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPICCTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFRSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFRSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
```

```

; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match
Best Local Similarity 99.8%; Score 1272; DB 14; Length 1132;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 953

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSADPNKEDHNVILGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHHYSTQSALSADPNKEDHNVILGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match
Best Local Similarity 99.2%; Score 1264; DB 9; Length 239;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSADPNKEDHNVILGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSADPNKEDHNVILGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match
Best Local Similarity 99.0%; Score 1261; DB 9; Length 239;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
```

```

; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match
Best Local Similarity 99.2%; Score 1264; DB 12; Length 239;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSADPNKEDHNVILGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYSTQSALSADPNKEDHNVILGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match
Best Local Similarity 99.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
```

QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAAMPEGVQVQERTIFFKDDGNKYNKTRAEVKFEGDYL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAAMPEGVQVQERTIFFKDDGNKYNKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHLSQSALSADPNKRDHMLVIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLSQSALSADPNKRDHMLVLEFVTAAGITLGMDELYK 239
RESULT 9
US-09-999-745-4
; Sequence 4, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-999-745-4

Query Match 99.0%; Score 1261; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
QY 51 LVTTLTSGVQCFSRYPDHMKQHDFFKSAAMPEGVQVQERTIFFKDDGNKYNKTRAEVKFEGDYL 120
DB 51 LVTTLTGVQCFSRYPDHMKQHDFFKSAAMPEGVQVQERTIFFKDDGNKYNKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNNHLSQSALSADPNKRDHMLVIGFVTAAGITLGMDELYK 239
DB 131 DHYQONTPIGDGPVLLPDNNHLSQSALSADPNKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 10
US-09-866-538-4
; Sequence 4, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: TSJEN, Roger
; APPLICANT: CAMPBELL, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT

; ORGANISM: Aequorea victoria
US-09-866-538-4
Query Match 99.0%; Score 1261; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAAMPEGVQVQERTIFFKDDGNKYNKTRAEVKFEGDYL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAAMPEGVQVQERTIFFKDDGNKYNKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHLSQSALSADPNKRDHMLVIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLSQSALSADPNKRDHMLVLEFVTAAGITLGMDELYK 239
RESULT 11
US-09-797-496B-2
; Sequence 2, Application US/09797496B
; Publication No. US20030049597A1
; GENERAL INFORMATION:
; APPLICANT: Simon, Sanford M.
; APPLICANT: Chen, Yu
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof
; FILE REFERENCE: 600-1-267
; CURRENT APPLICATION NUMBER: US/09/797,496B
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described
; OTHER INFORMATION: in specification
US-09-797-496B-2

Query Match 99.0%; Score 1261; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAAMPEGVQVQERTIFFKDDGNKYNKTRAEVKFEGDYL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAAMPEGVQVQERTIFFKDDGNKYNKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHLSQSALSADPNKRDHMLVIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLSQSALSADPNKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 12
US-09-794-308-4
; Sequence 4, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA

APPLICANT: TSIEH, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match 99.0%; Score 1261; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
Sequence 4, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEH, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REGEN1550
CURRENT APPLICATION NUMBER: US/09/865,291
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match 99.0%; Score 1261; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239
RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US20030212265A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Acsuehi
APPLICANT: Llopis, Juan
APPLICANT: Wächter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR APPLICATION NUMBER: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 99.0%; Score 1261; DB 12; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYVKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239
RESULT 15
US-10-121-258-13
Sequence 13, Application US/10121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP1
CURRENT APPLICATION NUMBER: US/10/121,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NC 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      99.0%; Score 1261; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 MVSKEBELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKLTLPKFICTTGKLPVPWPT 60
Db      1 MVSKEBELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKLTLPKFICTTGKLPVPWPT 60

QY      61 LVTTLSTGVQCFSRYPDPHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db      61 LVTTLSTGVQCFSRYPDPHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQALSXPNEKRDHVVLLIGFVTAAGITLGMDELYK 239
Db      181 DHYQONTPIGDGPVLLPDNHYLSTQALSXPNEKRDHVVLLIGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:28
Job time : 34.7778 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds
(without alignments)
2224.817 Million cell updates/sec

Title: US-09-887-784-222i
Perfect score: 1274
Sequence: 1 MVSKEELFTGVVPIVLVD.....VLIGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	97.3	238	1 JQ1514	green-fluorescent
2	104	8.2	785	2 H72228	hypothetical prote
3	91.5	7.2	861	2 H64102	leucine-trna ligas
4	90	7.1	632	2 T06586	DNA-binding protei
5	89.5	7.0	887	2 E82590	leucyl-trna synth
6	88.5	6.9	655	2 D83917	DNA topoisomerase
7	88.5	6.9	797	2 JC4078	protective surface
8	88.5	6.9	808	2 F64102	protective surface
9	87.5	6.9	370	2 E70390	iron-sulfur cofact
10	87.5	6.9	860	2 AC0582	leucyl-trna synth
11	87.5	6.9	941	2 S29043	cellulase (EC 3.2.
12	87.5	6.9	2222	1 A36028	DNA-directed DNA p
13	87.5	6.9	2573	2 D71614	hypothetical prote
14	87	6.8	578	1 I40794	dihydrolipoamide d
15	86	6.8	357	2 G81355	tRNA (uracil-5-) m
16	85.5	6.7	788	1 JCVLHH	DNA-directed DNA p
17	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
18	85	6.7	281	2 AD2052	hypothetical prote
19	85	6.7	874	2 JC9330	S-layer protein pr
20	84.5	6.6	425	2 C97354	hypothetical prote
21	84.5	6.6	613	2 A99552	oligoendopeptidase
22	84	6.6	353	2 E84941	imidazoleglycerol-
23	84	6.6	461	2 T06936	photosystem II chl
24	83.5	6.6	836	1 JDVLDD	DNA-directed DNA p
25	83.5	6.6	1134	2 A60234	Iga Fc receptor pr
26	83.5	6.6	1164	1 FCSOAG	Iga Fc receptor pr
27	83	6.5	439	2 JH0414	synaptogamin o-p65
28	82.5	6.5	534	1 NICLMA	nitrogenase (EC 1.
29	82.5	6.5	740	2 G95153	neuraminidase, pro

RESULT 1

JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C>Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:1347277
A:Accession: JQ1514
A:Molecule type: DNA
A:Residues: 1-107, 'S', 109-238 <PRA1>
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 'O', 26-156, 'P', 158-171, 'K', 173-238 <INO>
A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384
R:Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009
A:Experimental source: clone gfp1
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24, 'O', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 209-238 <PRA5>
A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011
A:Experimental source: clone gfp2
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65692; PDB:1GFL
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-95
A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58953; MUID:98294543; PMID:9631087

hypothetical prote
hypothetical prote
carboxy-terminal p
proprotein convert
alpha-amyase homo
lipoxigenase (EC 1
water-stress-induc
synergohymenotropi
hypothetical prote
hypothetical prote
hypothetical prote
leucine-trna ligas
leucine trna synth
leucine trna synth
hypothetical prote

A:Contents: annotation; X-ray crystallography, 1.9 angstroms
C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
C:Genetics:

A:Gene: GFP
A:Introns: 69/3; 167/3
A:Superfamily: green-fluorescent protein
C:Keywords: chromoprotein; luminescence
F:55-67/Cross-link: 5-Imidazolinone (Ser-Gly) #status experimental
F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.3%; Score 1239; DB 1; Length 238;
Best Local Similarity 96.6%; Pred. No. 3.2e-96;
Matches 230; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGHSVSGEGDATYKLTLPKFTCTTGKLPVPWPTL 61

Db 1 MSKGEELFTGVVPILVELDGVNKGHSVSGEGDATYKLTLPKFTCTTGKLPVPWPTL 60

QY 62 VTTLGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIPFKDDGNKYKTRAEVKFEGDTLV 121

Db 61 VTTFSGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIPFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181

Db 121 NRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLIGFTVTAAGITLGMDELYK 239

Db 181 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLIGFTVTAAGITLGMDELYK 238

RESULT 2

H72228
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72228
R: Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID: 99287316; PMID: 10360571

A:Accession: H72228

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-785 <ARN>

A:Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1624

Query Match 8.2%; Score 104; DB 2; Length 785;
Best Local Similarity 19.7%; Pred. No. 1.1;
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPILVELDGVNKGHSVSGEGDATYKLTLPKFTCTTGKLPVPWPTLV 62

Db 15 NEGRFSFEGTVPGVQAD-----LVRKGLLPHPYVGM- 46

QY 63 TTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIPFKDDGNKYKTRAEVKFEGDTLV 122

Db 47 -----NEDLFKEIDREWIYEREFKEDVKGERVDLVPFEGVDTLV 88

QY 123 RIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQKNGIKVNFIRHNIEDGSVOLADH 182

Db 89 DVTYLVGYL---GSTEDMFTEYFDTNVL-----KEKNHLKVYIK-----SPIRVKPT 134

QY 183 YQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLIGFTVTAAGITLGM 235

Db 135 LEQNYGLGGP-----EDP-----IRGYIRKAQYSYGM 163

RESULT 3

H64102
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)

N:Alternate names: leucyl-tRNA synthetase

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002

C:Accession: H64102

R: Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
Gocayne, J.D.; Scott, J.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.;
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C.

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID: 95350630; PMID: 7542800

A:Accession: H64102

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-861 <TIGR>

C:Genetics:

A:Gene: leuS

C:Superfamily: leucine-tRNA ligase

C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.2%; Score 91.5; DB 2; Length 861;
Best Local Similarity 24.1%; Pred. No. 14;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVPWPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIPFKD----- 103

Db 314 TGDKLPV-I-WVNFVLMHYGTGAVMAVPAH-DQDQF-----EFAQKYSLPKQVIAPLA 364

QY 104 DGNKYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNVSHNVIMADK-QKNGI 162

Db 365 DEEIDLTKQAFVEHGHKLVNSDEFQGNF--DGAFFN-----IADKLEKLG 408

QY 163 ---KNFKIRH-----NIEDGSVOLADHYQQWTPIGDGPVLLPDNHYL- 202

Db 409 GKRQNVNRDGNVSRQRYGAPIMPLTLENGDVPA-----PMEDLPILLPEDVMD 461

QY 203 STQSALSKDPN 213

Db 462 GVKSPINADPN 472

RESULT 4

T06586

DNA-binding protein PD2 - garden pea

C:Species: Pisum sativum (garden pea)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999

C:Accession: T06586

R:Sato, N.; Kazuno, A.A.; Ohta, N.; Ohshima, K.

submitted to the EMBL Data Library, June 1996

A:Description: Identification of a novel family of DNA-binding proteins with two AT-hook

A:Reference number: Z15774

A:Accession: T06586

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-632 <SAT>

A:Cross-references: EMBL:X98740; NID:e995229; PIDN:CAA67292.1; PID:e275185

A:Experimental source: cv. Alaska

Query Match 7.1%; Score 90; DB 2; Length 632;
Best Local Similarity 23.3%; Pred. No. 12;
Matches 49; Conservative 26; Mismatches 79; Indels 56; Gaps 7;

QY 16 LVLELDGVNKGHSVSGEGDATYKLTLPKFTCTTGKLPVPWPTLVTTLSYGVQCFSRY 75

Db 363 IVDQGRDVGSKVDVINKSNEATIPENK----PTEPKLDVEQELAAITM----- 408

QY 76 PDHMKQHDFFKSAMPEGYV-----QERTIFFKDDGNKYKTRAEVKFEGDTLVNR 125

```
Db 409 PSSAKNVNLTDLIVETFPPLRSVARTSSGREGSEELKDSGNSLERTDKKLELEQGNKS-E 467
QY 126 LKGIKDFKEDGNILGHKLE-----YNYGNHNVYIMADKQK-----N 160
Db 468 LKGIPTDSTLDEKFNALGNKILKEISNPRHDVESANHSTHNKQVTVSHQKALETNN 527
QY 161 GIKVNFKIRHINBDG-----SVQLADHYQ 184
Db 528 QSOVEDVAKNIQDDSKPSBESLHKDKVR 557

RESULT 5
E82590
leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: E82590
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82590
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-887 <SIM>
A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.P.; Marino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Teuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
A:Contents: annotation
C:Genetics:
A:Gene: XF2176
C:Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 887;
Best Local Similarity 22.2%; Pred. No. 21;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPWPTLVTLLSYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNY-- 107
Db 329 TNEQLPV-WVANFVLMVGTGAVMVPVGHQDQDEF--ANKYGLPRTQVIALEKPNKQDE 385
QY 108 -----KTRAEVKFEGLTLVNRIELKGIDFKEDGNILGHKLEYNNYNSHNVI 153
Db 386 STWEPDVRDWYADKTR--EFE--LINSAPFDGLDYQDAFEVLAERFE----- 429
QY 154 MADKQXNG-INKNFKIRHINIEDSGVQLADHYQNTPI-----GGCPVLLPDN 199
Db 430 ---ROGRQRRVYRLR-----DWGVSQRVYWGCFIPVIYCPTGCGAVPVPEQQLPVILPEN 482
QY 200 -HYLSQTQSALSADPNKSR 216
Db 483 VAFSGTGSPKTPDPEWRK 500

RESULT 6
D83917
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: D83917
```

```
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirar
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: D83917
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <STO>
A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05859.1; GSPDB:GN001
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2140
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 6.9%; Score 88.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK---FSVSGEGEGDAT---YGLTTLKPI-----CTTGKLPVPWP 59
Db 63 NVTIHKDQSVSVRDEGRGMPTGMHKLKGTPEVILTVLHAGGKFGGGVATSGGLHGVGA 122
QY 60 TLVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQER-----TIFKDDG----- 105
Db 123 SVVNALSEWLIVEIKRDGVVYEQRFENGKGPSTTLEKKGKTRGTGTHFKPDPTVFSTT 182
QY 106 --NYKTRAEVKFEGLTLVNRIELKGIDFKEDGNILGHKLEYNNYNSHNVIIMADK----- 157
Db 183 NFNVTLSRLREAAFLKGLKIELVDLRDDTKVEPH-YEDGKAFVEYLNEDKETLHPV 241
QY 158 -----QKNGIKVNFKIRHINIEDSGVQLADHYQNTPIGDPVLLPDNHYLSTQSALSADPN 212
Db 242 VFFNGESNGIEIEFAQFQFN--DGYTENVLFSVNNVATKDG-----GTHELGAKTAMTRAV 294
QY 213 NE 214
Db 295 NE 296

RESULT 7
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
A:Variety: type b
C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
R:Flack, F.S.; Loomore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl
A:Reference number: JC4078; MUID:95255676; PMID:7737523
A:Accession: JC4078
A:Molecule type: DNA
A:Residues: 1-797 <FLA>
A:Cross-references: GB:U39361; NID:9537447; PIDN:AAA85645.1; PID:G537448
A:Experimental source: type b
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 6.9%; Score 88.5; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. No. 22;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY 65 LSYGVQCFSRYPDHMKQHDFF-----FKSAMPEGYVQE-----RTI 99
Db 427 IGYGTESGISYQASVKQDNFLGTGAASVIAGTNDYGTGTSVNLGYTETPFYTKDGVSLGGNV 486
QY 100 FFKDDGNKYTRAEVKEFGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNNYS 148
Db 487 FFENYDNSKSDTSNNYKRTTYGNSVTL-GFPVNNNSYVVGLTGHTYNNKLSNFALETN--- 542
QY 149 HNVYIMADKQK-NGIKVNFKIRHINIEDSGVQLADHYQO-----NTPIGDGPVLL 196
```

```
Db      543 RNLIYQSMKFKNGIKTN-----DFDFSGFWNNSLNRRGYPFKGVKASLG-GRVTI 593
QY      197 P--DNHYLSTQSALSKDPNKRDMHVLIGFVTAAGITLG 233
Db      594 PGSDNKYYKLSADVQGFYPLDRDHLWVWSAKASAGYANG 632

RESULT 8
F64102
proteactive surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C:Accession: F64102
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: F64102
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-808 <TIGR>
A:Cross-references: GB:L42023; TIGR:HI0917
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen

Query Match      6.9%; Score 88.5; DB 2; Length 808;
Best Local Similarity 21.9%; Pred. No. 23;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY      65 LSVGVQCFSRYPDHMKQHDF-----FKSAMPEGYVQOE-----RTI 99
Db      440 IGYGTESGISYQASVKQDNFLGTGCAAVSIAGTKNDYGTSLNLTGTEPYFTKDGVSIGSNV 499
QY      100 FFKDDGNYKTRAEVKEGDTLVNRIELKIGDFKEDGNI---LGH-----KLENNYNS 148
Db      500 FFENYNSKSDTSNSRTYTGNSVTL-GFPVNNNSYVGLGHTYNTKSNFALEYN--- 555
QY      149 HNYVIMADKQK-NGIKVNFKIRHNIEDSGVQLADHYQQ-----NTPIGDGPVLL 196
Db      556 RNLIYQSMKFKNGIKTN-----DFDFSGFWNNSLNRRGYPFKGVKASLG-GRVTI 606
QY      197 P--DNHYLSTQSALSKDPNKRDMHVLIGFVTAAGITLG 233
Db      607 PGSDNKYYKLSADVQGFYPLDRDHLWVWSAKASAGYANG 645

RESULT 9
E70390
iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus
N;Contains: L-cysteine sulfoxidase (EC 2.8.1.-)
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999
C:Accession: E70390
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98156666; PMID:9537320
A:Accession: E70390
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-370 <AQF>
A:Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: nifs1
C:Superfamily: nitrogen fixation protein nifs
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfotransferase
F;195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted
```

```
F;318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match      6.9%; Score 87.5; DB 2; Length 370;
Best Local Similarity 25.4%; Pred. No. 9.9;
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY      4 KGEELFTGVV---PILVELD---GDVNGHKF-SVSGEG---EGDATVYKLTLLKFICT 50
Db      164 KGVPLLTDAVQAIGKPIELKNISVATFSGHFKHAIKGSGFLYISDEANVEPLIVGGQE 223
QY      51 TGKLP-----VPMPTLTVTLSYGVQCFSRYPDHMKQ-HDFFKSAMPEGYVQERTIFFKOD 104
Db      224 NKGSGTENVVGLSLAKALEIVSNFSYQQLKRLDLFENLLLEA-LPDAQIVGKDA 282
QY      105 GNYKTRAEV---KPEGDTLVNRIELKIGDFKEDGNTLGHKLEYNNSHNVYIMADKQKNG 161
Db      283 ERSPSISVIMPKFFGAETVNNKSEKGIYCSGACLSGEYBPNKMLKMGFSQEKALRM 342
QY      162 IKVNFKIRHNIED 174
Db      343 VRFSFGLLNKEE 355

RESULT 10
AC0582
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (st
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0582
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0582
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-860 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176
C:Genetics:
A:Gene: STY0699
C:Superfamily: leucine-tRNA ligase

Query Match      6.9%; Score 87.5; DB 2; Length 860;
Best Local Similarity 23.3%; Pred. No. 30;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY      50 TTGKLPVPMPTLTVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109
Db      314 TGEIIPV-WAANFVLMYEYGTGAVMVGH-DQRD-YEFASKYGLTIKPVILLADGSEPD 370
QY      110 RAEVFEGETLVNRIELKIGDFKEDGNTLGHKLEYNNSHNVYIMADKQKNGIKVNFKIR 169
Db      371 SEQALTEKGVLFNSGEFGDLAEAFNAIDKL-----AEKGVGERKVNRLR 418
QY      170 H-----NTEDSGVQLADHYQNTPIGDGPVLLPDNHYL-STQSALSQDP 212
Db      419 DWGVSQRQYWGAPIPMVLTLEDGT-----LPTPEDQLPVLIPEDVVMVDGITSPIKADP 471

RESULT 11
S29043
cellulase (EC 3.2.1.4) - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
C:Accession: S29043; PC4404
R;Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
J. Gen. Microbiol. 136, 1327-1334, 1990
A:Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from
```


Db 184 NLHIKENKKDEKKNKHNNNDNNM-----IYYKNI---DKTHYILDNNVVHILNDIN 236

QY 210 KDPNEKPDHM 219 :||:|

Db 237 TYLKRERDYM 246 :||:|

RESULT 14

I40794

dihydroliipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum

N;Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase

hydrogenase complex chain E3; S-complex 50K chain

C;Species: Clostridium magnum

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C;Accession: I40794

R;Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuechel, A.

J. Bacteriol. 176, 3614-3630, 1994

A;Title: Biochemical and molecular characterization of the Clostridium magnum acetoin de

A;Reference number: I40789; MUID:94266715; PMID:8206840

A;Accession: I40794

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-578 <KRU>

A;Cross-references: GB:L31844; NID:G472324; PIDN:AAA21748.1; PID:G472330

C;Function:

A;Description: catalyzes the oxidation of dihydroliipoamide to liipoamide using NAD

A;Pathway: acetoin dehydrogenase enzyme system

C;Superfamily: Alcaligenes dihydroliipoamide dehydrogenase; dihydroliipoamide dehydrogenase

C;Keywords: FAD; flavoprotein; liipoamide; NAD; oxidoreductase; redox-active disulfide

F;5-77/Domain: lipoyl/biotin-binding homology <LPS>

F;117-145/Region: beta-alpha-beta FAD nucleotide-binding fold

F;119-561/Domain: dihydroliipoamide dehydrogenase homology <DLD>

F;287-315/Region: beta-alpha-beta NAD nucleotide-binding fold

F;153-158/Disulfide bonds: redox-active #status Predicted

Query Match 6.8%; Score 87; DB 1; Length 578;

Best Local Similarity 23.5%; Pred. No. 19;

Matches 54; Conservative 40; Mismatches 84; Indels 52; Gaps 12;

QY 10 TGVPVILVELDGVNGHKFSVSGEGSDATYKGLTLK-----FICTTGKLPVPWFTLVT 63

Db 255 TGSMPFIPEIE-----GNKLS----GVIDST-GALSLESNPESIAITGGGVIGVEFASIFN 305

QY 54 TLSYGVQCFSRYPDHMKQHDFFKSPMEGYSQVQERTIFFKDDGNKYKRAEVKPEGDTLVNR 123

Db 306 SLGCKVSIIEMLPHLPMDREISEI-----AKALIRDGINNN 346

QY 124 IELKGIQDFKEDG---NILGHKLEYNVSHNVYIMADKQKN--GIKYNFKIRHNIEDGSVQ 178

Db 347 CKVTRIEQEGDGLKVSFIGDKGESIDVEKVLIAVGRRSNIEGLDVE-KIGVKTEGGSII 405

QY 179 LADHYQONT----PIGD--GPVLLPDNHYLSTQSALSKD-----PNEKRDH 218

Db 406 VNDKMETNVEGIYAGDCTGKIMLA--HVASDQGVVAENIMGQNKKMDY 453

RESULT 15

G81355

tRNA (uracil-5-)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter jejuni

C;Species: Campylobacter jejuni

C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C;Accession: G81355

R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell

Nature 403, 665-668, 2000

A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A;Reference number: A81250; MUID:20150912; PMID:10688204

A;Accession: G81355

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-357 <PAR>

A;Cross-references: GB:ALJ39076; GB:ALJ11168; NID:G6968128; PIDN:CAB73096.1; PID:G696827

A;Experimental source: serotype O2, strain NCTC 11168

C;Genetics:

A;Gene: trmA; Cj0831c

C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 6.8%; Score 86; DB 2; Length 357;

Best Local Similarity 24.8%; Pred. No. 13;

Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;

QY 80 KOHDFPKSAMPDGYVOERTIFFKDDGNKYKRAEVKF--EGDTLV-----NRIELKG 128

Db 14 EKHSFIKTYFKGFYTKDFKLFAKDKHYRTRAELSFYHENDTLFVAMFDPKSKKKYIIIEY 73

QY 129 IDPKED-----GNILGHKLEYNVSHNVYIMADKQKNKGIKYNFKIRHNIE 173

Db 74 LDFADEKICAFMRRLLEYLRQDNKLEKL-----FGVEFLTTKQE--LSITLLYHKNIIE 125

QY 174 D 174

Db 126 D 126

Search completed: June 21, 2004, 16:01:58

Job time : 11.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-222I

Perfect score: 1274

Sequence: 1 MVSKGEELFTGVVPIVLVELD.....VLIGFTVTRAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1245	97.7	238	1	GFP_AEQVI
2	91.5	7.2	861	1	SYL_HAEN
3	89.5	7.0	879	1	SYL_XYLFA
4	88.5	6.9	795	1	D152_HAEN
5	88.5	6.9	797	1	D151_HAEN
6	87.5	6.9	793	1	D153_HAEN
7	87.5	6.9	860	1	SYL_SALTI
8	87.5	6.9	860	1	SYL_SALTY
9	87.5	6.9	941	1	GUN_BACS6
10	87.5	6.9	2222	1	DPOE_YEAST
11	87	6.8	689	1	AC21_HUMAN
12	86.5	6.8	533	1	CP51_CANGA
13	86.5	6.8	879	1	SYL_XYLFT
14	86.5	6.8	1603	1	VIT4_CAEEL
15	86	6.8	357	1	TRMA_CAMJE
16	85.5	6.7	788	1	DPOI_HBHE
17	85.5	6.7	886	1	ITH3_MSAU
18	85	6.7	874	1	SLAP_BACLI
19	84.5	6.6	501	1	AMPA_WIGBR
20	84.5	6.6	613	1	PEPF_MYCPU
21	84.5	6.6	859	1	SYL_SHON
22	84	6.6	353	1	HIS7_BUCAI
23	84	6.6	366	1	SET1_HUMAN
24	84	6.6	461	1	PSBC_CYPAP
25	84	6.6	504	1	YC03_KLEPN
26	83.5	6.6	538	1	GRBE_RAT
27	83.5	6.6	1164	1	BAG_STRAG
28	83	6.5	439	1	SY62_DISOM
29	82.5	6.5	533	1	NIFP_CLOPA
30	82	6.4	682	1	NEC1_ECOLI
31	82	6.4	752	1	PRCI_RAT
32	82	6.4	774	1	AMY2_SCHPO
33	81.5	6.4	599	1	SYD_HAEDU

34	81.5	6.4	658	1	ADAS_HUMAN	O00116 homo sapien
35	81	6.4	682	1	AC2L_MOUSE	Q99nb1 mus musculus
36	80.5	6.3	860	1	SYL_ECO57	Q8xbn8 escherichia
37	80.5	6.3	860	1	SYL_ECOL6	Q8fjy9 escherichia
38	80.5	6.3	860	1	SYL_ECOLI	P07813 escherichia
39	80	6.3	336	1	YD48_METJA	Q08743 methanococc
40	80	6.3	461	1	PSBC_CHLEU	Q08684 chlamydomon
41	80	6.3	737	1	OPT1_DROME	P91679 drosophila
42	79.5	6.2	312	1	TRXB_CHLMU	Q9pkt7 chlamydia m
43	79.5	6.2	393	1	TRNB_HELPY	Q25443 helicobacte
44	79.5	6.2	468	1	GLNA_AZOCA	P94126 azorhizobiu
45	79.5	6.2	546	1	AMVB_BACCE	P36924 bacillus ce

ALIGNMENTS

RESULT 1						
GFP_AEQVI		STANDARD;		PRT;	238 AA.	
AC P42212; Q17104; Q27903;						
DT 01-NOV-1995 (Rel. 32, Created)						
DT 01-NOV-1995 (Rel. 32, Last sequence update)						
DT 10-OCT-2003 (Rel. 42, Last annotation update)						
DE Green fluorescent protein.						
GN GFP.						
OS Aequorea victoria (Jellyfish).						
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;						
OC Aequoreidae; Aequorea.						
OX NCBI_TaxID=6100;						
RN [1]						
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.						
RX MEDLINE=92175527; PubMed=1347277;						
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,						
RA Cormier M.J.;						
RT "Primary structure of the Aequorea victoria green-fluorescent						
RT protein.";						
RL Gene 111:229-233(1992).						
RN [2]						
RP SEQUENCE FROM N.A.						
RX MEDLINE=94185810; PubMed=8137953;						
RA Inouye S., Tsuji F.I.;						
RT "Aequorea green fluorescent protein. Expression of the gene and						
RT fluorescence characteristics of the recombinant protein.";						
RL FEBS Lett. 341:277-280(1994).						
RN [3]						
RP SEQUENCE FROM N.A.						
RX MEDLINE=97299832; PubMed=9154981;						
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;						
RT "Enhanced expression in tobacco of the gene encoding green fluorescent						
RT protein by modification of its codon usage.";						
RL Plant Mol. Biol. 33:989-999(1997).						
RN [4]						
RP CHROMOPHORE.						
RX MEDLINE=93192221; PubMed=8448132;						
RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;						
RT "Chemical structure of the hexapeptide chromophore of the Aequorea						
RT green-fluorescent protein.";						
RL Biochemistry 32:1212-1218(1993).						
RN [5]						
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).						
RX MEDLINE=96355665; PubMed=8703075;						
RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,						
RA Remington S.J.;						
RT "Crystal structure of the Aequorea victoria green fluorescent						
RT protein.";						
RL Science 273:1392-1395(1996).						
RN [6]						
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).						
RX MEDLINE=98294543; PubMed=9631087;						
RA Yang F., Moss L.G., Phillips G.N. Jr.;						
RT "The molecular structure of green fluorescent protein.";						
RL Nat. Biotechnol. 14:1246-1251(1996).						


```
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
RESULT 2
SYL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (Leurs).
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Karlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley K., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32774; AAC22581.1; -.
CC PIR; H64102; H64102.
CC TIGR; H10921; -.
CC HAMAP; MF_00049; -.
CC InterPro; IPR002302; Leu-trNASynt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001432; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_fiers_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMS; TIGR00396; leuS bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FE7 CRC64;
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 7;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVPTLVTLSYGVCFSRYPDHMKQHDFFKSAPEGYQVERIFFKD----- 103
DB 314 TGDKLPI-WYANFVLMHYGTGMVAFPAH-DQRDF-----EFAQKYSIPIKQVIAPLA 364
```

```
QY 104 DGNKYTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNNYNNHVIYIMADK-QKNGI 162
DB 365 DBEIDLTKQAFVEHKGKLVNSDFDGKNF--DGAFFNG-----IADKLEKLGV 408
QY 163 ---KVNFKIRH-----NIETDSVOLADHYQQNTPIGDPVLLPDNHYL- 202
DB 409 GKRQVNYRLRDGWSRQRYWGAPIPMLTLENGDVVFA-----PMEDLPILPEDVVMD 461
QY 203 STQSALSQDPN 213
DB 462 GYKSPINADPN 472
RESULT 3
SYL_XYLFA
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9PEG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (Leurs).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000)
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004031; AAF84975.1; ALT_INIT.
CC HAMAP; MF_00049; -.
DR
```



```
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AL627267; CAD05125.1; -.
DR EMBL; AE016841; AAC069822.1; -.
DR HAMAP; MF 00049; -.
DR InterPro; IPR002302; Leu-TRNAsynt1a.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009008; ValRS_1IERS_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 860 AA; 96940 MW; 2F95E480BBAB23C4 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 15;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVPWPTLVTTLSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKT 109
Db 314 TGEIIPV-WAANFVMEYGTGAVWVPGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKPEGDTLVNRIELKGIDFKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFKIR 169
Db 371 SEQALTEKGVLFNSGFEFDGLAFEAFAFNAIADKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NIEDGSVOLADHYQNTPIGDGPVLLPDNHYL-STQSALS KDP 212
Db 419 DWGVSQRQYWGAPIPWVLTLEDGTV-----LPTPEDQLPVILPDVMDGITSPKADP 471

RESULT 8
SYL_SALTY STANDARD; PRT; 860 AA.
AC Q8ZQ26;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR STM0648.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2."
RL Nature 413:852-856 (2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M27420; AAA22304.1; -.
DR PIR; S23043; S29043.
DR PDB; 1G01; 31-DEC-02.
```

```
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE008725; AAL19599.1; -.
DR StyGene; SG????; leus.
DR HAMAP; MF 00049; -.
DR InterPro; IPR002302; Leu-TRNAsynt1a.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009008; ValRS_1IERS_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 860 AA; 96985 MW; D5003584DFCCAB6 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 15;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVPWPTLVTTLSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKT 109
Db 314 TGEIIPV-WAANFVMEYGTGAVWVPGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKPEGDTLVNRIELKGIDFKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFKIR 169
Db 371 SEQALTEKGVLFNSGFEFDGLAFEAFAFNAIADKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NIEDGSVOLADHYQNTPIGDGPVLLPDNHYL-STQSALS KDP 212
Db 419 DWGVSQRQYWGAPIPWVLTLEDGTV-----LPTPEDQLPVILPDVMDGITSPKADP 471

RESULT 9
GUN_BACS6 STANDARD; PRT; 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Alkaline cellulase).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=2230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT "Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from Bacillus sp. KSM-635."
RL J. Gen. Microbiol. 136:1327-1334 (1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 3 s-layer homology (SLH) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M27420; AAA22304.1; -.
DR PIR; S23043; S29043.
DR PDB; 1G01; 31-DEC-02.
```

```
DR PDB; 1GOC; 31-DEC-02.
DR InterPro; IPR005086; CBM_17_28.
DR InterPro; IPR008979; Gal_bind like.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR001119; SLH.
DR Pfam; PF03424; CBM_17_28; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat;
KW 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 941 ENDOGLUCANASE.
FT DOMAIN 40 99 SLH 1.
FT DOMAIN 100 151 SLH 2.
FT DOMAIN 152 225 SLH 3.
FT ACT_SITE 373 373 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 485 485 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 941 AA; 104628 MW; BEA2AC3B169BFADA CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 941;
Best Local Similarity 20.3%; Pred. No. 16;
Matches 49; Conservative 33; Mismatches 62; Indels 97; Gaps 11;

QY 16 LVELDGVNGHKFSVSGEGDGYGKLTLCFTTCKLPVPMPTLTTLTSLVGVQCFSRY 75
D 16 LVELDGVNGHKFSVSGEGDGYGKLTLCFTTCKLPVPMPTLTTLTSLVGVQCFSRY 75
D 240 LVELNG-----QUTLAGE---DGT-----PVQLRGWST---HGLQWFG-- 271
QY 76 PDHMKQHDFFKSMGPGYQVQERTIFFKDDGNYKTRAEVKEGDTLVNRIELKGIDPKDG 135
D 76 PDHMKQHDFFKSMGPGYQVQERTIFFKDDGNYKTRAEVKEGDTLVNRIELKGIDPKDG 135
D 272 -ETVNEAFVSLNDGWSNIRLAMYIGENYATNPEVK---DLVYEGIELA----- 319
QY 136 NILGHKLEYNNSHNYIMADKQNGIKVNFKRHNIEDGSV---QLADHYQONTPIGD 191
D 136 NILGHKLEYNNSHNYIMADKQNGIKVNFKRHNIEDGSV---QLADHYQONTPIGD 191
D 320 -----FHDYVIYVDWH---VHAPGDPADVSGAYDFFBEIADHYKDH----- 360
QY 192 GPVLLPDNHYLSQTSALSKDPN-----EKRDHVMVLI 222
D 192 GPVLLPDNHYLSQTSALSKDPN-----EKRDHVMVLI 222
D 361 -----PNHYIWEIANEFPNNGGGLTNDKGEAVKEAYEPIVEMLRKGDNMILV 415
QY 223 G 223
D 416 G 416

RESULT 10
DPOE_YEAST STANDARD; PRT; 2222 AA.
AC P21951;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA polymerase II subunit A).
GN POL2 OR DUN2 OR YNL262W OR N0825.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.
RX MEDLINE=90381771; PubMed=2169349;
RA Morrison A., Araki H., Clark A.B., Hanatake R.K., Sugino A.;
RT "A third essential DNA polymerase in S. cerevisiae.";
RL Cell 62:1143-1151(1990).
RN [2]
RP SEQUENCE OF 1-2221 FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Flegl U., Niedenthal R.K., Hegemann J.H.;
RT "The sequence of a 24,152 bp segment from the left arm of chromosome XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
```


Db 473 FFGIVPVLMDKGSV-----VEGSNVSGALCIS-----QAWPGMARTI--- 510
QY 69 VQCFSRYPDMKQHDFFKSAMPEGYQERTIFFKDDGNKYKTRA---EVKPEGDTLVNRIE 125
Db 511 -----YGDHQRVDAYFRAYP-GY-----YFTGDGARTEGGYQITGRMDVVI----- 553
QY 126 LKGIDPKEDGNILGHKL 142
Db 554 -----NISGRL 560

RESULT 12
CP51_CANGA STANDARD; PRT; 533 AA.
AC PS0859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIA1) (Sterol 14-alpha-demethylase) (lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E., Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell viability, cell growth, sterol composition, and antifungal susceptibility.";
RT Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J., Rossler M.;
RT "Rapid detection and identification of Candida albicans and Torulopsis (Candida) glabrata in clinical specimens by species-specific nested PCR amplification of a cytochrome P-450 lanosterol-alpha-demethylase (lial) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is critical for ergosterol biosynthesis. It transforms lanosterol into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By similarity).
CC -!- CATALYTIC ACTIVITY: Obtusifolliol + 3 O(2) + 3 NADPH = 4-alpha-methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3 NADP(+) + 3 H(2)O.
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L40389; AAB02329.1; -;
CC EMBL; S75389; AAB32679.1; -;
CC InterPro; IPRO01128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
KW Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).

FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.4%; Pred. No. 10;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;
QY 25 GHKFSVS---GSGEGDATYKGLTKFKICTGKLPVWPMTLVTLTSLGVQCFSRYPDH--M 79
Db 109 GHEFTFNKALDVSAEAAYSHL-----TTPVFGKGVYDCPNHRLM 149
QY 80 KQHDFKFSAM-PEGYV-----QERTIFFKDDGNKYKTRAEVKFEKGGDTLVNRIELKGI 131
Db 150 EQKKFVKGALTKEAFVRYVPLIAEEIYKYFRSNKFNENNSGIVDVVMVQPEM--TIF 207
QY 132 KEDGNILGHKLEYNVSHNVYIMADKQKGIKVNFKIRNIHEDSGVOLADHYQONTPIGD 191
Db 208 TASRSLLGKEMRDKLDTDFAYLYSLDKLGFTTFINF--VFPNLPLEHYRKRDHAQQAIS--- 263
QY 192 GPVLLPDNHYLSTQSALS KDPNEKRD 217
Db 264 -----GYMSLIKERREKND 278

RESULT 13
SYL_XYLFT STANDARD; PRT; 879 AA.
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR PDI230.
OS Xylella fastidiosa (strain Temecual / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B., Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H., Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R., Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M., Carrier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J., Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E., Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M., Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V., da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T., Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D., de Souza A.A., Truffi D., Teukumo F., Yanai G.M., Zaros L.G., Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C., Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of Xylella fastidiosa.";
RL J. Bacteriol. 185:1018-1026(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP + diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE012557; AAO29080.1; ALT_INIT.
DR HAMAP; MF_00049; -; 1.


```
DR InterPro; IPR002302; Leu-TRNAsynt1a.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009008; ValRS_fiers_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA TRNA_LICASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99823 MW; 4C2EE01B8FDC497E CRC64;

Query Match
Best Local Similarity 22.2%; Pred. No. 18;
Matches 44; Conservative 28; Mismatches 69; Indels 57; Gaps 10;

QY 50 TTCKLVPWPTLVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDCGNY-- 107
DB 321 TNEQLPV-WVANFVLMAYGTGAVNVPFGHDQRQEF--ANKYGLPIRQVIALKEPKNQDE 377

QY 108 -----KTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNYSNNVYI 153
DB 378 SIWEPDVRDWDYADKTR---EFE--LINSAEFDGLDYQGA FEVLAEFE----- 421

QY 154 MADKQKNG-IKVNFKIRHNIEDGSVQLADHYQONTPI-----GDGPVLLPDN 199
DB 422 ---RQGRGRRVNYRLR---DMGVSQRQYWGCFPIVYCTGCAVPVPENQLPVILPEN 474

QY 200 -HYLSTQSALS KDPNEKR 216
DB 475 VAFSGTGSPKTPDEWRK 492

RESULT 14
VIT4_CABEL STANDARD; PRT; 1603 AA.
AC PI8947; Q9BPP3;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
GN VIT-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T., Spieth J., Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=85269643; PubMed=4022780;
RA Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;
RT "The C. elegans vitellogenin genes: short sequence repeats in the
RT promoter regions and homology to the vertebrate genes."
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -1- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
CC building the intestine of adult hermaphroditic individuals; they
CC are cotranslationally secreted into the body cavity and
CC subsequently taken up by the gonad.
CC -1- SIMILARITY: Contains 1 VWFD domain.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC024137; AAK09074.1; -.
CC DR EMBL; M11498; AAK28163.1; -.
CC DR EMBL; X02754; CAA26531.1; -.
CC DR PIR; A43084; A43084.
CC WormPep; F59D8.2; CE26817.
CC InterPro; IPR001747; Lipid_transprt_N.
CC InterPro; IPR001846; VWF_D.
CC Pfam; PF01347; Vitellogenin_N; 1.
CC SMART; SM00638; LFD_N; 1.
CC SMART; SM00216; VWD; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1603 VWFD.
FT DOMAIN 1308 1455
FT CONFLICT 30 30 Y -> V (IN REF. 3).
FT CONFLICT 169 169 L -> V (IN REF. 2).
FT CONFLICT 183 187 EVAYT -> RSLRH (IN REF. 2).
FT CONFLICT 275 275 T -> S (IN REF. 2).
SQ SEQUENCE 1603 AA; 186307 MW; E303170325BC99BB CRC64;

Query Match
Best Local Similarity 6.8%; Score 86.5; DB 1; Length 1603;
Matches 52; Conservative 32; Mismatches 69; Indels 69; Gaps 12;

QY 1 MYSKGEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60
DB 162 MESDKDSLFFNVHEKTMGDCVE---AYTIVQEG-GKTYTKSVNFDKCIIR-----PE 211

QY 61 LVTTLSTYGVQCFSRYPDHMKQHDFFKSAPEG-YVQERTIF---FKDDG----- 105
DB 212 TAYGLRFGSEC-----KECKEGQFVQPIVYTFYFKNEKLQSEVNSIYT 257

QY 106 -----NYKTRAVKPEGDTLVNRIELKIDFKEDGNILGHKLEYNYSNNVIMAD 156
DB 258 LNVNGQEVVVKSETRAKVTVEESKINR-EIK-----KVGPKKEIIVYSMENEKLIQ 308

QY 157 KQKNG-----IKVNFKIRHNIEDGSVQLADHYQONT 188
DB 309 FYKQGDKAENVNPPKALIEIQKV-EQLEEIFRQIQEH-EQNTTP 348

RESULT 15
TRMA_CAMJE STANDARD; PRT; 357 AA.
AC Q9PP92;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-
DE methyltransferase) (RUMT).
GN TRMA OR CJ0831C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds
(without alignments)
2458.984 Million cell updates/sec

Title: US-09-887-784-222I

Perfect score: 1274

Sequence: 1 MVSKGEELFTGVVPIVLVELD.....VLIGFVTRAGITLGMDELKY 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: *

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1241	97.4	238	2	Q8GHE2
2	1238	97.2	238	5	Q93125
3	1236	97.0	238	2	Q8GHE4
4	1235	96.9	238	2	Q8GHE3
5	1203	94.4	238	5	Q17105
6	1188	93.2	238	5	Q17106
7	1083	85.0	238	5	Q8WTC6
8	1079	84.7	238	5	Q8WPC95
9	1075	84.4	238	5	Q8WTC4
10	1073	84.2	238	5	Q8WTC5
11	1072	84.1	238	5	Q8WTC8
12	1072	84.1	238	5	Q8WTC9
13	1070	84.0	238	5	Q8WTC7
14	1068	83.8	238	5	Q8WTC5
15	252.5	19.8	225	5	Q95UA7
16	252.5	19.8	225	5	Q720W5

17	247	19.4	225	5	Q963F5
18	244.5	19.2	236	5	Q8T6U0
19	242.5	19.0	225	5	Q720W9
20	240	18.8	225	5	Q8I6J8
21	238.5	18.7	266	5	Q9U6Y3
22	233	18.3	225	5	Q720W4
23	232	18.2	224	5	Q8MU48
24	232	18.2	225	5	Q8T5F1
25	214	16.8	227	5	Q720W6
26	214	16.8	234	5	Q720W7
27	214	16.8	239	5	Q8MMA1
28	212.5	16.7	239	5	Q8MMA2
29	209	16.4	234	5	Q8T5F2
30	209	16.4	234	5	Q8MU47
31	208.5	16.4	229	5	Q9U6Y6
32	206	16.2	227	5	Q962P9
33	206	16.2	227	5	Q720W8
34	205.5	16.1	232	5	Q9GP15
35	205.5	16.1	238	5	Q9BLV9
36	204.5	16.1	214	5	Q86LV7
37	204	16.0	221	5	Q95P04
38	203.5	16.0	214	5	Q86LV8
39	202	15.9	227	5	Q95VT0
40	201.5	15.8	225	5	Q9U6Y8
41	201.5	15.8	232	5	Q9GZ28
42	198.5	15.6	222	5	Q72168
43	198.5	15.6	225	5	Q8T6T9
44	197.5	15.5	232	5	Q9U6Y7
45	197	15.5	235	5	Q8T5F0

ALIGNMENTS

RESULT 1

Q8GHE2 ID Q8GHE2 PRELIMINARY; PRT; 238 AA.
AC Q8GHE2;
DT 01-MAR-2003 (TREMREL. 23, Created)
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Green fluorescence protein.
GN 2289GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=DSM2289;
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324408; AAN86140.1; -
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;

Query Match 97.4%; Score 1241; DB 2; Length 238;
Best Local Similarity 97.5%; Pred. No. 3.6e-96;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY	2	VSKGEELFTGVVPIVLVELDGVNGHKFVSVGEGDATYKGLTKLFCITGKLPVPWPTL	61
Db	1	MKSGELFTGVVPIVLVELDGVNGHKFVSVGEGDATYKGLTKLFCITGKLPVPWPTL	60
QY	62	VTTLSVGVOCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNVYKTRAEVKEGDTLV	121

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF324405; AAN86137.1; -;
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP_like.
DR InterPro: IPR00786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;
Query Match 97.0%; Score 1236; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 9.6e-96;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 2 VSKGELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTKLFCITGKLPVPWPTL 61
DB 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTKLFCITGKLPVPWPTL 60
QY 62 VTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGTDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 238
RESULT 4
Q8GHE3 PRELIMINARY; PRT; 238 AA.
ID Q8GHE3
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF324406; AAN86138.1; -;
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP_like.
DR InterPro: IPR00786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP; 1.
DR ProDom: PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;
Query Match 96.9%; Score 1235; DB 2; Length 238;

Db 61 VTTFSGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 238
RESULT 2
Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-Optimized mutants of the green fluorescent protein (GFP).";
RL Microbiology 0:0-0(1996).
DR EMBL: U73901; AAB18957.1; -;
DR HSP; P42212; IBFP.
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP_like.
DR InterPro: IPR00786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;
Query Match 97.2%; Score 1238; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 6.5e-96;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 2 VSKGELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTKLFCITGKLPVPWPTL 61
Db 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTKLFCITGKLPVPWPTL 60
QY 62 VTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Db 61 VTTFSGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 238
RESULT 3
Q8GHE4 PRELIMINARY; PRT; 238 AA.
ID Q8GHE4
AC Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)

```
Best Local Similarity 97.1%; Pred. No. 1.2e-95;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLPICTTGKLPVWPPTL 61
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLPICTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db VTTFSYGVCFSRYPDHMKRHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRNIEDGSVOLAD 181
Db NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLIGFVTAAGITLGMDELYK 239
Db HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58789.1; -.
DR PIR; J50692; JQ1514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.4%; Score 1203; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 5.6e-93;
Matches 223; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLPICTTGKLPVWPPTL 61
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGVVPIILVELDGVNGQKFSVSGEGDATYGLTKLPICTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRNIEDGSVOLAD 181
Db NRIELKGIDFKEDGNILGHKMEYNYNHNVYIMGDKPKNGIKVNFKIRNIKDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLIGFVTAAGITLGMDELYK 239
Db HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAARITHGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
```

```
Q17106;
01-NOV-1996 (TrEMBLrel. 01, Created)
01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; J50692; JQ1514.
DR HSSP; P42212; 1BPP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;

Query Match 93.2%; Score 1188; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 1e-91;
Matches 221; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLPICTTGKLPVWPPTL 61
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGVVPIILVELDGVNGQKFSVSGEGDATYGLTKLPICTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRNIEDGSVOLAD 181
Db NRIELKGIDFKEDGNILGHKMEYNYNHNVYIMGDKPKNGIKVNFKIRNIKDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLIGFVTAAGITLGMDELYK 239
Db HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTSAGITHGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.O., Chen M.,
RA Li S.J., Xia N.S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
```

DR PRINTS; PRO1229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 6B8FD75E89926903 CRC64;

Query Match 85.0%; Score 1083; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 6.6e-83;
Matches 195; Conservative 21; Mismatches 22; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHKFSVSGEGDGYKLTLPKFTCTTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPIVLVELDGVNKHKFSVSGEGDGYKLTLPKFTCTTGKLPVWPPTL 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 62 VTTLISYGVQCSRYPDHMKOHDFKSAPEGYVOERTIFFKDDGNKYNKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTFYSYGIQCFARYPEHMKMNDFFKSAPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 122 NRLEKGDIDFKEDGNILGHKLEYNVNSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRLEKGMDFKEDGNILGHKLEYNFNSHNYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVIGFVTAAGITLGMDELYK 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLIPIINHLYSTQTAISKDRNETRDHMLVLEFFSACGHTGMDLYK 238
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
ID Q8WP95
AC Q8WP95;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP STRAIN=GFPxm;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qian Y.X., Pang S.Q.,
RA Li S.J., Xia N.S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013824; AAK02062.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27049 MW; 8185D05E529012B CRC64;

Query Match 84.7%; Score 1079; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 1.4e-82;
Matches 194; Conservative 21; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHKFSVSGEGDGYKLTLPKFTCTTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPIVLVELDGVNKHKFSVSGEGDGYKLTLPKFTCTTGKLPVWPPTL 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 62 VTTLISYGVQCSRYPDHMKOHDFKSAPEGYVOERTIFFKDDGNKYNKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTFYSYGIQCFARYPEHMKMNDFFKSAPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 122 NRLEKGDIDFKEDGNILGHKLEYNVNSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRLEKGMDFKEDGNILGHKLEYNFNSHNYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVIGFVTAAGITLGMDELYK 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLIPIINHLYSTQTAISKDRNETRDHMLVLEFFSACGHTGMDLYK 238
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 9
Q8WTC4 PRELIMINARY; PRT; 238 AA.
ID Q8WTC4
AC Q8WTC4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ShG24;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RL "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match 84.4%; Score 1075; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 3.1e-82;
Matches 196; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHKFSVSGEGDGYKLTLPKFTCTTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPIVLVELDGVNKHKFSVSGEGDGYKLTLPKFTCTTGKLPVWPPTL 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 62 VTTLISYGVQCSRYPDHMKOHDFKSAPEGYVOERTIFFKDDGNKYNKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLISYGVQCSRYPDHMKOHDFKSAPEGYVOERTIFFKDDGNKYNKTRAEVKFEGDTLV 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 122 NRLEKGDIDFKEDGNILGHKLEYNVNSHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRLEKGMDFKEDGNILGHKLEYNFNSHNYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVIGFVTAAGITLGMDELYK 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLIPIINHLYSTQTAISKDRNETRDHMLVLEFFSACGHTGMDLYK 238
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
ID Q8WTD0
AC Q8WTD0;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RL "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds
(without alignments)
1433.395 Million cell updates/sec

Title: US-09-887-784-222L

Perfect score: 1274
Sequence: 1 MYSKGEELFTGVVILVELD.....VLGLFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:.*
1: Genesecp1980s:.*
2: Genesecp1990s:.*
3: Genesecp2000s:.*
4: Genesecp2001s:.*
5: Genesecp2002s:.*
6: Genesecp2003as:.*
7: Genesecp2003bs:.*
8: Genesecp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1274	100.0	239	5 AAE17518	Aae17518 Enhanced
2	1274	100.0	363	6 ABR40352	Abr40352 Human ami
3	1274	100.0	893	4 AAG65781	Aag65781 Amino aci
4	1274	100.0	1132	4 AAG55782	Aag55782 Amino aci
5	1266	99.4	239	5 AAE17517	Aae17517 Enhanced
6	1263	99.1	239	3 AAB22882	Aab22882 Enhanced
7	1263	99.1	239	3 AAY54349	Aay54349 Amino aci
8	1263	99.1	239	3 AAY79584	Aay79584 EGFP sign
9	1263	99.1	239	4 AAB50804	Aab50804 Jellyfish
10	1263	99.1	239	4 AAB58590	Aab58590 A. victor
11	1263	99.1	239	4 AAB31171	Aab31171 Amino aci
12	1263	99.1	239	5 AAG66198	Aag66198 A. victor
13	1263	99.1	239	5 AAG94444	Aag94444 Protease
14	1263	99.1	239	5 AAE14599	Aae14599 Aequorea
15	1263	99.1	239	6 AAE34958	Aae34958 Aequorea
16	1263	99.1	239	6 AAG79829	Aag79829 Green flu
17	1263	99.1	239	6 ABR83616	Abr83616 Green flu
18	1263	99.1	239	6 ADA38074	Ada38074 Aequorea
19	1263	99.1	239	7 ABUE63204	Abue63204 Aequorea
20	1263	99.1	239	7 ADC18358	Adc18358 EGFP (enh
21	1263	99.1	239	7 ABW00914	Abw00914 Aequorea
22	1263	99.1	239	7 ADE28570	Ade28570 Enhanced
23	1263	99.1	246	7 ABM79011	Abm79011 Enhanced
24	1263	99.1	248	5 AAG68319	Aag68319 Jellyfish
25	1263	99.1	259	5 AAU99804	Aau99804 Biomembra

26	1263	99.1	265	2 AAW97451	Aaw97451 Wild-type
27	1263	99.1	268	5 AAU99803	Aau99803 Biomembra
28	1263	99.1	270	5 AAU99802	Aau99802 Biomembra
29	1263	99.1	272	5 AAU99800	Aau99800 Biomembra
30	1263	99.1	273	5 AAU99801	Aau99801 Biomembra
31	1263	99.1	280	5 AAU99807	Aau99807 Biomembra
32	1263	99.1	281	3 AAY50142	Aay50142 Green flu
33	1263	99.1	281	3 AAB24252	Aab24252 EGFP-MODC
34	1263	99.1	281	5 AAU10888	Aau10888 EGFP-MODC
35	1263	99.1	286	7 ADE28562	Ade28562 EGFP/ hum
36	1263	99.1	289	7 ADE28564	Ade28564 EGFP/ hum
37	1263	99.1	290	7 ADE28568	Ade28568 EGFP/ hum
38	1263	99.1	290	7 ADE28566	Ade28566 EGFP/ hum
39	1263	99.1	294	3 AAB22860	Aab22860 GFP-DSVD-
40	1263	99.1	294	3 AAY79638	Aay79638 Caespase-3
41	1263	99.1	294	5 ABG94422	Abg94422 Recombina
42	1263	99.1	308	2 AAY42181	Aay42181 EGFP/DRM
43	1263	99.1	320	6 ABR83620	Abr83620 HUB1-GFP
44	1263	99.1	323	3 AAY54359	Aay54359 GFP mutan
45	1263	99.1	323	6 ABR83621	Abr83621 RUB1-GFP

ALIGNMENTS

RESULT 1
AAE17518
ID AAE17518 standard; protein; 239 AA.
XX AAE17518;
AC AAE17518;
DT 22-APR-2002 (first entry)
XX
DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.
XX
KW Jellyfish; green fluorescent protein; GFP; protein redistribution;
KW cellular function; genetic reporter; mutant; Stoke's shift; mutain.
XX Aequorea victoria.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 65 /note= "Wild type Phe substituted with Leu; This
FT corresponds to position 64 in the wild type protein"
FT
FT Misc-difference 223 /note= "Wild type Glu substituted with Gly; This
FT corresponds to position 222 in the wild type protein"
FT
FT WO2001983338-A2.
PD 27-DEC-2001.
XX
PF 18-JUN-2001; 2001WO-EF006848.
XX
PR 19-JUN-2000; 2000DK-00000953.
PR 20-JUN-2000; 2000US-0212681P.
PR 10-MAY-2001; 2001BK-00000739.
PR 10-MAY-2001; 2001US-0290170P.
XX
PA (BIOI-) BIOIMAGE AS.
XX
PI Bjorn SP, Pagliaro L, Thastrup O;
XX
DR WPI; 2002-098224/13.
DR N-PSDB; AAD28163.
XX
PT Novel fluorescent protein in in vitro assay for measuring protein kinase
PT activity or dephosphorylation activity, or for measuring protein
PT redistribution, has a green fluorescent protein with F64L and E222G
PT mutation.
XX
PS Claim 9; Page 37; 41pp; English.

XX The invention relates to a fluorescent protein derived from green
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
CC at F64L and E222G has a bigger compared to other GFP's making it very
CC suitable for high throughput screening due to better resolution. The
CC fluorescent protein is useful in *in vitro* assays for measuring protein
CC kinase activity or dephosphorylation activity, or for measuring protein
CC redistribution. The fluorescent protein is useful in studying cellular
CC functions in living cells; as protein tags in transgenic animals, living
CC and fixed cells; organelle tags, secretion marker and genetic integrity.
CC The fluorescent protein is also useful as a cell or organelle integrity
CC marker, a marker for changes in cell morphology, as transfection marker,
CC and as a marker to be used in combination with fluorescence activated
CC cell sorting (FACS). The novel proteins can also be used as reporters to
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
CC protein is also useful as markers in transcriptional and translational
CC fusions for performing transposon vector mutagenesis and as a reporter
CC for bacterial detection. Transposons encoding the fluorescent protein are
CC useful for screening promoters and for tagging plasmids and chromosomes.
CC The fluorescent protein engineered into the genome of a phage is useful
CC for designing diagnostic tool. The present sequence is a DNA encoding
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
XX
SQ Sequence 239 AA;
Query Match 100.0%; Score 1274; DB 5; Length 239;
Best Local Similarity 100.0%; Pred. No. 4.8e-123;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRLEKIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 2
ABR40352
ID ABR40352 standard; protein; 363 AA.
XX
AC ABR40352;
XX
DT 08-JUL-2003 (first entry)
XX
DE Human amino acid sequence SEQ ID NO: 6.
XX
KW Human; heterologous conjugate; intracellular protein.
XX
OS Homo sapiens.
OS Aequoria victoria.
XX
FN WO2003029827-A2.
XX
PD 10-APR-2003.
XX
PF 01-OCT-2002; 2002WO-DK000651.
XX
XX 01-OCT-2001; 2001DK-00001433.
PR 11-OCT-2001; 2001US-0328868P.
XX
XX (BIOL-) BIOLMAGE AS.
XX
XX Terry BR, Nielsen SJ;

XX WPI: 2003-430211/40.
DR N-PSDB; ACC72604.
XX
PT Novel cell for identifying modulators of protein interaction, contains a
PT first conjugate comprising anchor protein, second conjugate having type B
PT interactor protein and third conjugate with detectable group.
XX
PS Disclosure; Page 112-113; 118pp; English.
XX
CC The invention relates to a novel cell, comprising three heterologous
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that
CC specifically binds to an internal structure within the cell conjugated to
CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
CC type B conjugated to a first protein of interest, and a third HC (HC3)
CC comprising a second protein of interest conjugated to detectable group.
CC The cell is useful for detecting if a compound disrupts or induces the
CC interaction between two intracellular proteins. The cell is also useful
CC for screening compounds that modulate the interaction between two
CC intracellular proteins. The present sequence is used in the
CC exemplification of the invention
XX
SQ Sequence 363 AA;
Query Match 100.0%; Score 1274; DB 6; Length 363;
Best Local Similarity 100.0%; Pred. No. 8.8e-123;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRLEKIDPFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 3
AAG65781
ID AAG65781 standard; protein; 893 AA.
XX
AC AAG65781;
XX
DT 07-JAN-2002 (first entry)
XX
DE Amino acid sequence of HSPDE4A1-E222G fusion protein.
XX
KW PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW fusion protein.
XX
OS Homo sapiens.
OS Aequorea victoria.
XX
FN WO200179526-A2.
XX
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-DK000264.
XX
XX 17-APR-2000; 2000DK-00000651.
PR 29-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX

PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
DR WPI: 2001-611727/70.
DR N-PSDB; AA166852.
XX
PT Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
PS Example 1; Page 156-160; 160pp; English.
XX
CC The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxin shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
CC protein
XX
SQ Sequence 893 AA;
Query Match 100.0%; Score 1274; DB 4; Length 893;
Best Local Similarity 100.0%; Pred. No. 3.3e-122; Indels 0; Gaps 0;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
DB 655 MYSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 893
RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA.
AC AAG65782;
XX
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX fusion protein.
XX Homo sapiens.
OS Aequorea victoria.
XX
PN WO200179526-A2.
XX

PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-DK000264.
XX
PR 17-APR-2000; 2000DK-00000651.
PR 29-MAY-2000; 2000DK-00000849.
XX
PA (BIOI-) BIOIMAGE AS.
XX
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
XX WPI: 2001-611727/70.
DR N-PSDB; AA166853.
XX
PT Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
PS Example 1; Page 162-167; 160pp; English.
XX
CC The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxin shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
CC protein
XX
SQ Sequence 1132 AA;
Query Match 100.0%; Score 1274; DB 4; Length 1132;
Best Local Similarity 100.0%; Pred. No. 4.7e-122;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
DB 894 MYSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 953
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 1132
RESULT 5
AAE17517
ID AAE17517 standard; protein; 239 AA.
XX
AC AAE17517;
XX
XX AAE17517;
DT 22-APR-2002 (first entry)
XX
XX Enhanced F64L jellyfish green fluorescent protein mutant.
DE
XX

KW	Jellyfish; green fluorescent protein; GFP; protein redistribution;	
KW	cellular function; genetic reporter; mutant; Stoke's shift; muten.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 65	
FT	/note= "Wild type Phe substituted with Leu; This	
FT	corresponds to position 64 in the wild type protein"	
FT		
PN	WO200198338-A2.	
XX		
PD	27-DEC-2001.	
XX		
PF	18-JUN-2001; 2001WO-EP006848.	
XX		
PR	19-JUN-2000; 2000DK-00000953.	
PR	20-JUN-2000; 2000US-0212681P.	
PR	10-MAY-2001; 2001DK-00000739.	
PR	10-MAY-2001; 2001US-0290170P.	
XX		
PA	(BIOI-) BIOIMAGE AS.	
XX		
PI	Bjorn SP, Pagliaro L, Thastrup O;	
XX		
DR	WPI; 2002-098224/13.	
DR	N-PSDB; AAD28162.	
XX		
PT	Novel fluorescent protein in in vitro assay for measuring protein kinase	
PT	activity or dephosphorylation activity, or for measuring protein	
PT	redistribution, has a green fluorescent protein with F64L and E222G	
PT	mutation.	
XX		
PS	Example 1; Page 35; 41pp; English.	
XX		
CC	The invention relates to a fluorescent protein derived from green	
CC	fluorescent protein (GFP) or its analogue. The GFP containing mutations	
CC	at F64L and E222G has a bigger compared to other GFP's making it very	
CC	suitable for high throughput screening due to better resolution. The	
CC	fluorescent protein is useful in invitro assays for measuring protein	
CC	kinase activity or dephosphorylation activity, or for measuring protein	
CC	redistribution. The fluorescent protein is useful in studying cellular	
CC	functions in living cells; as protein tags in transgenic animals, living	
CC	and fixed cells; organelle tags, secretion marker and genetic reporter.	
CC	The fluorescent protein is also useful as a cell or organelle integrity	
CC	marker, a marker for changes in cell morphology, as transfection marker,	
CC	and as a marker to be used in combination with fluorescence activated	
CC	cell sorting (FACS). The novel proteins can also be used as reporters to	
CC	monitor live or dead biomass of organisms, such as fungi. The fluorescent	
CC	protein is also useful as markers in transcriptional and translational	
CC	fusions for performing transposon vector mutagenesis and as a reporter	
CC	for bacterial detection. Transposons encoding the fluorescent protein are	
CC	useful for screening promoters and for tagging plasmids and chromosomes.	
CC	The fluorescent protein engineered into the genome of a phage is useful	
CC	for designing diagnostic tool. The present sequence is enhanced F64L	
CC	Jellyfish green fluorescent protein (GFP) mutant	
XX		
SQ	Sequence 239 AA;	
	Query Match 99.4%; Score 1266; DB 5; Length 239;	
	Best Local Similarity 99.6%; Pred. No. 3.2e-122;	
	Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 MVSNGEELFTGVPILVELDGVNGHKFSVSGEGEDATYKLTLPKICTTGKLPVWPWT 60	
DB	1 MVSNGEELFTGVPILVELDGVNGHKFSVSGEGEDATYKLTLPKICTTGKLPVWPWT 60	
QY	61 LVTTLSYGVQCFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNVKTAEVKFEGDTL 120	
DB	61 LVTTLSYGVQCFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNVKTAEVKFEGDTL 120	
QY	121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSVOLA 180	
DB	121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSVOLA 180	
QY	181 DHYQNTPIGDGPFVLLPDNHYLSTQSALSKDPEKRDHMLVLFVTAAGITLGMDELYK 239	
DB	181 DHYQNTPIGDGPFVLLPDNHYLSTQSALSKDPEKRDHMLVLFVTAAGITLGMDELYK 239	
QY	181 DHYQNTPIGDGPFVLLPDNHYLSTQSALSKDPEKRDHMLVLFVTAAGITLGMDELYK 239	
DB	181 DHYQNTPIGDGPFVLLPDNHYLSTQSALSKDPEKRDHMLVLFVTAAGITLGMDELYK 239	
RESULT 6		
AAB22882		
ID	AAB22882 standard; protein; 239 AA.	
XX		
AC	AAB22882;	
XX		
DT	10-JAN-2001 (first entry)	
XX		
DE	Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.	
XX		
KW	Bioreporter protein; fusion protein; recognition site;	
KW	cellular targeting sequence; cellular localisation; fluorescent protein;	
KW	protease activity detection; toxin detection; cellular stress detection;	
KW	drug discovery; cell based screening.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
PN	WO200050872-A2.	
XX		
PD	31-AUG-2000.	
XX		
PF	25-FEB-2000; 2000WO-US004794.	
XX		
PR	26-FEB-1999; 99US-0122152P.	
PR	08-MAR-1999; 99US-0123399P.	
PR	12-JUL-1999; 99US-00352171.	
XX		
PA	(CELL-) CELLOMICS INC.	
XX		
PI	Giuliano KA, Kapur R;	
XX		
DR	WPI; 2000-594086/56.	
DR	N-PSDB; AAA93373.	
XX		
PT	Automated cell-based characterization of toxin by contacting cells	
PT	containing luminescent reporter molecules with test substance and	
PT	analyzing optically.	
XX		
PS	Example 11; Fig 29A; 336pp; English.	
XX		
CC	The invention relates to systems, methods and reagents for cell-based	
CC	screening or detection of compounds which affect particular biological	
CC	functions. The methods of the invention utilise fluorescent bioreporter	
CC	molecules which, when acted on by a compound of interest, cause an	
CC	alteration in the cellular distribution of at least the fluorescent	
CC	moiety. In one embodiment, the biosensors comprise heat shock proteins	
CC	(HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent	
CC	protein (GFP), or derivatives thereof). Such biosensors are located in	
CC	the cytoplasm, but on stress activation translocate to the nucleus. In	
CC	another embodiment bioreporter proteins can be used to detect protease	
CC	activity. Such protease bioreporter proteins comprise one or more	
CC	fluorescent proteins; a recognition signal which is cleaved by the	
CC	protease; and at least one cellular localisation signal. The latter two	
CC	components may be components of a single protein which is acted upon by	
CC	the protease, or may be from heterologous sources. Due to the	
CC	localisation signal, the bioreporter protein is localised to a particular	
CC	region of the cell. Once acted on by the protease of interest, the	
CC	fluorescent protein is cleaved from the localisation sequence, and is	
CC	free to migrate to other locations within the cell. The presence of a	
CC	second localisation signal attached to the fluorescent protein enables	
CC	the fluorescent protein to be directed to a different cellular	
CC	compartment after cleavage of the protease recognition sequence. The	
CC	change in distribution of the fluorescent protein can be detected using	
CC	imaging methods with a high degree of spatial resolution. The methods and	

CC biosensors of the invention can be used to investigate a wide range of
CC cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased quantity
CC and speed of data collection and faster evaluation of drug candidates.
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used
CC as components of biosensor fusion proteins of the invention
XX
SQ Sequence 239 AA;

Query Match 99.1%; Score 1263; DB 3; Length 239;
Best Local Similarity 99.2%; Pred. No. 6.6e-122;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 7
AAV54349
ID AAY54349 standard; protein; 239 AA.
XX AAY54349;
XX
XX
DT 06-APR-2000 (first entry)
XX
DE Amino acid sequence of the mutant green fluorescent protein EGFP.
XX
XX Fluorescent protein; green fluorescent protein; emission intensity;
KW fluorescence; pH detection; pH sensor; EGFP.
XX
OS Synthetic.
OS Aequorea victoria.

XX
FH Key Location/Qualifiers
FT Misc-difference 65 /note= "wild type Phe substituted with Leu"
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"
FT Misc-difference 232 /note= "wild type His substituted with Leu"
XX
XX WO9964592-A2.
XX
XX 16-DEC-1999.
XX
XX 08-JUN-1999; 99WO-US012850.
XX
XX 09-JUN-1998; 98US-00094359.
XX 13-OCT-1998; 98US-00172063.
XX
XX (REGC) UNIV CALIFORNIA.
PA (UYOR-) UNIV OREGON STATE.
XX

PI Tsien RY, Llopis J, Wachter RM;
XX WPI; 2000-116540/10.
DR N-PSDB; AA245642.
XX
PT New functional engineered green fluorescent proteins, used for measuring
XX the pH in biological samples and cells.
XX
PS Disclosure; Page 9; 89pp; English.
XX
CC The present sequence represents a functional engineered fluorescent
CC protein based on the Aequorea green fluorescent protein (GFP). The
CC emission intensity changes as pH varies between 5 and 10 of the present
CC protein are novel. The functional engineered fluorescent proteins show
CC reversible changes in fluorescence over physiological pH ranges. They can
CC be used for determining the pH of samples and cells. The polynucleotides
CC can also be used to produce transgenic animals. The fluorescent protein
CC pH sensors can be delivered to cells in the form of polynucleotides
CC encoding the protein sensor fused to a targeting signal. The targeting
CC signal directs the expression of the protein sensors to restricted cell
CC locations. This makes it possible to measure the pH of a precisely
CC defined cellular region or organelle
XX
SQ Sequence 239 AA;

Query Match 99.1%; Score 1263; DB 3; Length 239;
Best Local Similarity 99.2%; Pred. No. 6.6e-122;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 8
AAV79584
ID AAY79584 standard; peptide; 239 AA.
XX AAY79584;
XX
XX
DT 29-AUG-2000 (first entry)
XX
DE EGFP signal domain.
XX
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;
KW analysis; drug discovery.
XX
OS Unidentified.
XX
XX WO200026408-A2.
XX
XX 11-MAY-2000.
XX
XX 29-OCT-1999; 99WO-US025431.
XX
XX 30-OCT-1998; 98US-0106308P.
XX 26-MAY-1999; 99US-0136078P.
XX
XX (CELL-) CELLOMICS INC.
XX
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX WPI; 2000-365644/31.
DR N-PSDB; AAA27573.
XX Recombinant nucleic acid encoding a protease biosensor useful for
PT fluorescence based cell and molecular biochemical assays for drug
PT discovery comprising three operably linked nucleic acid sequences.
XX
XX Claim 14; Fig 29A; 218pp; English.
XX
XX The present sequence is that of the EGFP signal domain, which can be
CC included in novel recombinant protease biosensors (PBs) of the invention.
CC The PBs (see AA79638-54) comprise: a first domain (see AA79579-87)
CC comprising at least 1 detectable polypeptide signal such as the present
CC sequence; a second domain (see AA79588-622) comprising at least 1
CC protease recognition site; and a third domain (see AA79623-37)
CC comprising at least 1 reactant target sequence. A recombinant nucleic
CC acid (see AAA27627-43) encoding the PB, an expression vector, and a
CC genetically engineered host cell are also claimed. A claimed method for
CC identifying compounds that modify protease activity in a cell involves
CC contacting a host cell that possesses the recombinant PB with a test
CC compound, and determining the PB distribution in the host cell, where
CC changes in the distribution of the PB are correlated with modification of
CC protease activity by the test compound. Claimed kits for identifying
CC compounds that modify protease activity in a host cell include the
CC recombinant nucleic acid, or the recombinant PB, or the vector, or the
CC host cell. The PB is useful in high content screens to detect in vivo
CC activation of enzymatic activity, and to identify specific activity based
CC on cleavage of a known recognition motif
XX
XX Sequence 239 AA;
SQ

Query Match 99.1%; Score 1263; DB 3; Length 239;
Best Local Similarity 99.2%; Pred. No. 6.6e-122;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 9
AAB50804
ID AAB50804 standard; protein; 239 AA.
XX
XX AAB50804;
AC
XX
DT 14-MAR-2001 (first entry)
XX
DE Jellyfish GFP mutant EGFP.
XX
KW Aequorea victoria; jellyfish; fluorescent protein indicator;
KW green fluorescent protein; GFP; linker moiety; sensor;
KW calmodulin-binding domain; mutant; mutein.
XX
OS Aequorea victoria.
XX
FN WO200071565-A2.
XX
PD 30-NOV-2000.
XX

PF 17-MAY-2000; 2000WO-US013684.
XX
PR 21-MAY-1999; 99US-00316919.
PR 21-MAY-1999; 99US-00316920.
XX
XX (REGC) UNIV CALIFORNIA.
PA
XX Tsien RY, Baird GA;
PI
XX WPI; 2001-032017/04.
DR N-PSDB; AAC90488.
XX
XX Novel fluorescent proteins comprising a sensor protein inserted into
PT them, useful for measuring the response of a sensor biological, chemical,
PT electrical or physiological parameter in vivo or in vitro.
XX
XX Disclosure; Page 24; 94pp; English.
PS
XX The present sequence is a fluorescent protein used in the construction of
CC a fluorescent protein indicator. The indicator comprises a sensor
CC polypeptide that is responsive to a chemical, biological, electrical or
CC physiological parameter, and a fluorescence protein functional group. The
CC sensor polypeptide is operatively inserted into the fluorescent moiety.
CC The fluorescent indicator is useful for detecting the presence of a
CC response inducing member in a sample. The method involves contacting the
CC sample with the indicator and detecting a change in fluorescence, in
CC which a change is indicative of the effect of the parameter on the sensor
CC polypeptide. The novel fluorescent proteins are advantageous due to their
CC reduced size as compared to the FRET (fluorescence resonance energy
CC transfer)-based sensors
XX
XX Sequence 239 AA;
SQ

Query Match 99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 6.6e-122;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 10
AAB85900
ID AAB85900 standard; protein; 239 AA.
XX
XX AAB85900;
AC
XX
DT 30-NOV-2001 (first entry)
XX
DE A. victoria green fluorescent protein (GFP) and linker sequence.
XX
KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
KW green fluorescent polypeptide; orexinergic; anabolic; food intake; GFP;
KW green fluorescent protein.
XX
OS Synthetic.
OS Aequorea victoria.
XX
FN WO200168706-A1.
XX

```

PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-US008071.
XX
PR 15-MAR-2000; 2000US-0189698P.
XX
PA (MERI ) MERCK & CO INC.
XX
PI Marsh DJ;
XX
PI WPI; 2001-565791/63.
XX
DR N-PSDB; AAH47304.
XX
XX Fusion proteins comprising melanin concentrating hormone receptor
PT peptides and fluorescent proteins, useful for identifying appetite
PT stimulants.
XX
PS Claim 2; Page 14; 71pp; English.
XX
CC The invention provides melanin concentrating hormone (MCH) receptor
CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
CC MCHR polypeptide regions from different species. The MCHR fusion protein
CC comprise MCHR polypeptide region and a fluorescent polypeptide region
CC joined directly, or via a linker, to the carboxy side of the MCHR
CC polypeptide region. The MCHR fusion proteins can be expressed by standard
CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
CC regulation of MCH activity stimulates food intake. The present sequence
CC represents a A. victoria green fluorescent protein (GFP) and a linker
CC sequence
XX
XX Sequence 239 AA;

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 6.6e-122;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 11
AAB31171
XX AAB31171 standard; protein; 239 AA.
XX
AC AAB31171;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of a green fluorescent protein (GFP).
XX
KW Growth rate; death rate; reporter gene; luminescent protein;
KW fluorescent product; luciferase; green fluorescent protein; GFP.
XX
OS Aequorea victoria.
XX
PN WO200075367-A1.
XX
PD 14-DEC-2000.
XX
PF 07-JUN-2000; 2000WO-FI000507.

XX
XX 07-JUN-1999; 99FI-00001296.
XX
XX (LILI/) LILIUS E.
PA (VIRT/) VIRT M.
XX
XX Lilius E, Virta M;
XX
XX WPI; 2001-061737/07.
DR N-PSDB; AAC86954.
XX
XX Assessing growth and death rates of a micro-organism in a desired
PT environment, by introducing 2 reporter genes encoding luminescent and
PT fluorescent products and detecting luminescent fluorescence.
XX
XX Disclosure; Page 27; 32pp; English.
XX
XX The specification describes a method for assessing the growth rate and
CC death rate of a micro-organism within a predetermined time period in a
CC desired environment. The method comprises introducing at least two
CC reporter genes encoding luminescent and/or fluorescent products into the
CC micro-organisms, incubating the micro-organism within the desired
CC environment, and detecting luminescence and/or fluorescence after a
CC predetermined time period. Use of two different markers within a micro-
CC organism enables the differentiation between growth and death rates. The
CC method is used to assess the growth rate and death rate of a micro-
CC organism within a predetermined time period in a desired environment. The
CC present sequence represents a green fluorescent protein (GFP), and is
CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
CC and is used in the method of the invention
XX
XX Sequence 239 AA;

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 6.6e-122;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 12
AAG66198
XX AAG66198 standard; protein; 239 AA.
XX
AC AAG66198;
XX
DT 17-JUN-2002 (first entry)
XX
DE A. victoria green fluorescent protein (EGFP).
XX
KW Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
KW green fluorescent protein; EGFP.
XX
OS Aequorea victoria.
XX
PN JP2002045189-A.
XX
PD 12-FEB-2002.
XX

```

PF	04-AUG-2000; 2000JP-00237165.	17-SEP-1999; 99US-00398965.	PR
XX		29-OCT-1999; 99US-00430656.	PR
PR	04-AUG-2000; 2000JP-00237165.	01-DEC-1999; 99US-0168408P.	PR
XX			XX
XX	(RIKA) RIKAGAKU KENKYUSHO.	(GIUL/) GIULIANO K.	PA
PA		(KAPU/) KAPUR R.	PA
XX			XX
DR	WPI; 2002-299190/34.	Giuliano K, Kapur R;	PI
DR	N-PSDE; ABL40628.		XX
XX		WPI; 2002-634730/68.	XX
PT	A gene encoding cyan-green fluorescent protein.	DR N-PSDB; ABS71491.	DR
XX			XX
PS	Example; Page 14; 20pp; Japanese.		PT
XX		Automated cell-based toxin detection, classification, and/or	PT
CC	The invention relates to a gene encoding proteins having cyan-green	identification by treating cells involves use of three classes of	PT
CC	fluorescence characteristic and having a function of showing stable	luminescent reporter molecules such as detectors, classifiers or	PT
CC	fluorescence characteristic in acid region. A method for the preparation	identifiers.	PT
CC	of a cyan-green fluorescent protein is provided which involves a		XX
CC	transformant transformed by a recombinant vector comprising the gene,	Example 10; Fig 29A; 214pp; English.	PS
CC	where the transformant is cultured and the protein is collected from the		XX
CC	culture. The present sequence represents the A. victoria green	The invention describes methods of automated detection, classification	CC
CC	fluorescent protein (EGFP)	and identification comprising treating cells containing luminescent	CC
XX		reporter molecules (I) in array of locations with a test substance, where	CC
SQ	Sequence 239 AA;	(I) are detectors, classifiers or identifiers, imaging cells in each	CC
		location to obtain luminescent signals and converting optical information	CC
		into digital data to interpret presence of toxins in the test substance.	CC
		The method are useful for detection of toxins chosen from proteases, ADP-	CC
		ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.	CC
		Three classes of cell-based luminescent reporter molecules such as	CC
		detectors, classifiers and identifiers are described and serve as	CC
		reporters of toxic threat agents. The first two levels of	CC
		characterisation ensure a rapid readout of toxin class without	CC
		sacrificing the ability to detect many new mutant toxins or dissect	CC
		several complex mixtures of known toxins. This is the amino acid sequence	CC
		of a protease biosensor related signal sequence used in the cell-based	CC
		screening system	CC
			XX
		Sequence 239 AA;	SQ
		Query Match 99.1%; Score 1263; DB 5; Length 239;	
		Best Local Similarity 99.2%; Pred. No. 6.6e-122;	
		Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	
Qy	1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTAFICTTGKLPVPWPT 60		Qy
Db	1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTAFICTTGKLPVPWPT 60		Db
Qy	61 LVTTLSYGVQCFGRYPDHPKQHDFFKSAPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120		Qy
Db	61 LVTTLSYGVQCFGRYPDHPKQHDFFKSAPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120		Db
Qy	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180		Qy
Db	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180		Db
Qy	181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239		Qy
Db	181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239		Db
		RESULT 14	
		AAE14599	
		ID AAE14599 standard; protein; 239 AA.	
		XX	
		XX	
		AC AAE14599;	
		XX	
		XX	
		DT 31-MAY-2002 (first entry)	
		XX	
		DE Aequorea victoria enhanced green fluorescent protein.	
		XX	
		XX	
		XX	
		XX	
		XX	
		OS Aequorea victoria.	
		OS Synthetic.	

XX Key Location/Qualifiers
FH Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"
FT
FT
FT
FT Misc-difference 65 /note= "GFP phe64 is replaced by Leu"
FT
FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"
FT
FT
PN EP1178109-A1.
XX
XX
PD 06-FEB-2002.
XX
PF 03-AUG-2001; 2001EP-00306650.
XX
PR 04-AUG-2000; 2000JP-00237166.
XX
XX (RIKE) RIKEN KK.
XX
XX Miyawaki A, Sawano A;
PI
PI
DR WPI; 2002-208112/27.
DR N-PSDB; AAD27910.
XX
XX
XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimer.
PS Example 1; Page 13-14; 31pp; English.
XX
XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimer. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimer, the megaprimer are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria
XX
SQ Sequence 239 AA;

Query Match 99.1%; Score 1263; DB 5; Length 239;
Best Local Similarity 99.2%; Pred. No. 6.6e-122;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKGELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60
Db 1 MVSKGELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDITL 120
Db 61 LVTTLTGYQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDITL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 15
AAE34958
ID AAE34958 standard; protein; 239 AA.
XX
AC AAE34958;
XX
DT 28-MAY-2003 (first entry)
XX
DE Aequorea victoria enhanced green fluorescent protein (EGFP).
XX
KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;
KW kinase; enhanced green fluorescent protein; EGFP.
XX
OS Aequorea victoria.
XX
PN WO200295058-A2.
XX
PD 28-NOV-2002.
XX
PF 24-MAY-2002; 2002WO-US016955.
XX
PR 24-MAY-2001; 2001US-00865291.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Ting AY, Zhang J;
XX
DR WPI; 2003-148474/14.
DR N-PSDB; AAD53428.
XX
PT Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphaminoacid binding domain, and acceptor molecule, in operative linkage.
XX
PS Disclosure; Col 56-57; 38pp; English.
XX
XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or operative linkage, a donor molecule, a phosphorylatable domain, a phosphaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention
XX
SQ Sequence 239 AA;

Query Match 99.1%; Score 1263; DB 6; Length 239;
Best Local Similarity 99.2%; Pred. No. 6.6e-122;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKGELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60
Db 1 MVSKGELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDITL 120
Db 61 LVTTLTGYQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDITL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

Db 181 DHYQQNTPIGDPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:11
Job time : 48.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-222L
Perfect score: 1274
Sequence: 1 MWSKGEELFTGVVPIVLVD.....VLGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2.6/prodata/2/iaa/5A COMB.pdp.*
2: /cgn2.6/prodata/2/iaa/5B COMB.pdp.*
3: /cgn2.6/prodata/2/iaa/6A COMB.pdp.*
4: /cgn2.6/prodata/2/iaa/6B COMB.pdp.*
5: /cgn2.6/prodata/2/iaa/PCTUS COMB.pdp.*
6: /cgn2.6/prodata/2/iaa/backfiles1.pdp.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1263	99.1	239	3	US-09-172-063-3
2	1263	99.1	239	4	US-09-513-783A-46
3	1263	99.1	239	4	US-09-316-919-4
4	1263	99.1	239	4	US-09-602-641-3
5	1263	99.1	239	4	US-09-920-922-2
6	1263	99.1	281	3	US-09-062-102-1
7	1263	99.1	281	4	US-09-364-946-1
8	1263	99.1	294	4	US-09-513-783A-2
9	1263	99.1	323	3	US-09-172-063-21
10	1263	99.1	323	4	US-09-602-641-21
11	1263	99.1	364	3	US-09-085-305-6
12	1263	99.1	379	4	US-09-417-197-129
13	1263	99.1	434	4	US-09-800-170-48
14	1263	99.1	442	4	US-09-417-197-127
15	1263	99.1	459	4	US-09-513-783A-170
16	1263	99.1	544	4	US-09-417-197-113
17	1263	99.1	544	4	US-09-417-197-115
18	1263	99.1	604	4	US-09-417-197-59
19	1263	99.1	605	4	US-09-417-197-41
20	1263	99.1	606	4	US-09-417-197-65
21	1263	99.1	607	4	US-09-417-197-47
22	1263	99.1	630	4	US-09-417-197-63
23	1263	99.1	631	4	US-09-417-197-39
24	1263	99.1	633	4	US-09-417-197-45
25	1263	99.1	635	4	US-09-417-197-125
26	1263	99.1	642	2	US-08-818-253-2
27	1263	99.1	642	2	US-08-818-253-6

28	1263	99.1	642	3	US-08-818-252-2	Sequence 2, Appli
29	1263	99.1	642	3	US-08-818-252-6	Sequence 6, Appli
30	1263	99.1	652	2	US-08-818-253-4	Sequence 4, Appli
31	1263	99.1	652	3	US-08-818-252-4	Sequence 4, Appli
32	1263	99.1	718	4	US-09-417-197-75	Sequence 75, Appli
33	1263	99.1	719	4	US-09-417-197-51	Sequence 51, Appli
34	1263	99.1	726	4	US-09-417-197-71	Sequence 71, Appli
35	1263	99.1	727	4	US-09-417-197-139	Sequence 139, App
36	1263	99.1	783	4	US-09-513-783A-176	Sequence 176, App
37	1263	99.1	797	4	US-09-417-197-141	Sequence 141, App
38	1263	99.1	797	4	US-09-417-197-143	Sequence 143, App
39	1263	99.1	798	4	US-09-417-197-77	Sequence 77, Appli
40	1263	99.1	805	4	US-09-513-783A-178	Sequence 178, App
41	1263	99.1	806	4	US-09-417-197-53	Sequence 53, Appli
42	1263	99.1	836	4	US-09-417-197-61	Sequence 61, Appli
43	1263	99.1	842	4	US-09-417-197-43	Sequence 43, Appli
44	1263	99.1	843	4	US-09-417-197-117	Sequence 117, App
45	1263	99.1	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1
US-09-172-063-3
; Sequence 3, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Lloplis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-172-063-3

Query Match	99.1%;	Score	1263;	DB	3;	Length	239;
Best Local Similarity	99.2%;	Pred. No.	1.4e-127;				
Matches	237;	Conservative	1;	Mismatches	1;	Indels	0;
QY	1	MWSKGEELFTGVVPIVLVDGVNKHKFSVSGEGDATYVGKLTAKFICTTGKLPVWPWT	60				
Db	1	MWSKGEELFTGVVPIVLVDGVNKHKFSVSGEGDATYVGKLTAKFICTTGKLPVWPWT	60				
QY	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNKYKTRAEVKFEGDTL	120				
Db	61	LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNKYKTRAEVKFEGDTL	120				
QY	121	VNRIELKGDIFKEDGNILGHKLEYNYSNNVIMADKQNGIKVNFKIRHNIEDGSVOLA	180				
Db	121	VNRIELKGDIFKEDGNILGHKLEYNYSNNVIMADKQNGIKVNFKIRHNIEDGSVOLA	180				
QY	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLLGFVTAAGITLGMDELYK	239				
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLLGFVTAAGITLGMDELYK	239				

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.4e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.4e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wächter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.4e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.4e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          99.1%; Score 1263; DB 3; Length 281;
Best Local Similarity 99.2%; Pred. No. 1.8e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CIP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          99.1%; Score 1263; DB 4; Length 281;
Best Local Similarity 99.2%; Pred. No. 1.8e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match          99.1%; Score 1263; DB 4; Length 294;
Best Local Similarity 99.2%; Pred. No. 2e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
|||
```

Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTCLKFICTTGKLPVWPWT 60
QY 61 LVTTLISYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQOLA 180
Db 121 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQOLA 180
QY 181 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9

US-09-172-063-21
; Sequence 21, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-172-063-21

Query Match 99.1%; Score 1263; DB 3; Length 323;
Best Local Similarity 99.2%; Pred. No. 2.3e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTCLKFICTTGKLPVWPWT 60
Db 85 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTCLKFICTTGKLPVWPWT 144
QY 61 LVTTLISYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 204
QY 121 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQOLA 180
Db 205 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQOLA 264
QY 181 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21
; Sequence 21, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-602-641-21
Query Match 99.1%; Score 1263; DB 4; Length 323;
Best Local Similarity 99.2%; Pred. No. 2.3e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTCLKFICTTGKLPVWPWT 60
Db 85 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTCLKFICTTGKLPVWPWT 144
QY 61 LVTTLISYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 145 LVTTLTYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 204
QY 121 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQOLA 180
Db 205 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQOLA 264
QY 181 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6
; Sequence 6, Application US/09085305
; Patent No. 6191269
; GENERAL INFORMATION:
; APPLICANT: Pollock, Allan
; APPLICANT: Lovett, David H.
; APPLICANT: Turck, Johanna
; TITLE OF INVENTION: Selective Induction of Apoptosis in
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESS: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,305
; FILING DATE: 29-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDGYGKLTLLKFICTTGKLPVWPMT 60
QY 61 LVTTLISYGVQCSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170

Query Match 99.1%; Score 1263; DB 4; Length 459;
Best Local Similarity 99.2%; Pred. No. 3 8e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDGYGKLTLLKFICTTGKLPVWPMT 60
Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDGYGKLTLLKFICTTGKLPVWPMT 60
QY 61 LVTTLISYGVQCSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:03
Job time : 13.7778 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds

(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-222L

Perfect score: 1274
Sequence: 1 MVSKEELFTGVVPIVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processor: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1274	100.0	239	9	US-09-887-784-4
2	1274	100.0	239	12	US-10-296-953-4
3	1274	100.0	363	14	US-10-270-223-6
4	1274	100.0	893	14	US-10-257-909A-30
5	1274	100.0	1132	14	US-10-257-909A-32
6	1266	99.4	239	9	US-09-887-784-2
7	1266	99.4	239	12	US-10-296-953-2
8	1263	99.1	239	9	US-09-920-922-2
9	1263	99.1	239	9	US-09-999-745-4
10	1263	99.1	239	10	US-09-866-538-4
11	1263	99.1	239	10	US-09-797-496B-2
12	1263	99.1	239	10	US-09-794-308-4
13	1263	99.1	239	10	US-09-865-291-4
14	1263	99.1	239	12	US-10-457-982-3
15	1263	99.1	239	14	US-10-121-258-13

16	1263	99.1	239	14	US-10-221-461-7	Sequence 7, Appli
17	1263	99.1	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1263	99.1	239	14	US-10-177-390-2	Sequence 2, Appli
19	1263	99.1	239	14	US-10-338-411-3	Sequence 3, Appli
20	1263	99.1	239	15	US-10-370-570-4	Sequence 4, Appli
21	1263	99.1	239	15	US-10-389-640-3	Sequence 3, Appli
22	1263	99.1	259	14	US-10-314-861-11	Sequence 11, Appli
23	1263	99.1	281	12	US-09-931-232-1	Sequence 1, Appli
24	1263	99.1	288	14	US-10-314-861-37	Sequence 37, Appli
25	1263	99.1	293	14	US-10-314-861-35	Sequence 35, Appli
26	1263	99.1	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1263	99.1	295	14	US-10-314-861-39	Sequence 39, Appli
28	1263	99.1	299	14	US-10-314-861-33	Sequence 33, Appli
29	1263	99.1	305	14	US-10-314-861-31	Sequence 31, Appli
30	1263	99.1	308	14	US-10-033-717-35	Sequence 35, Appli
31	1263	99.1	311	14	US-10-314-861-29	Sequence 29, Appli
32	1263	99.1	320	14	US-10-338-411-11	Sequence 11, Appli
33	1263	99.1	320	15	US-10-389-640-11	Sequence 11, Appli
34	1263	99.1	323	12	US-10-457-982-21	Sequence 21, Appli
35	1263	99.1	323	14	US-10-338-411-7	Sequence 7, Appli
36	1263	99.1	323	14	US-10-338-411-13	Sequence 13, Appli
37	1263	99.1	323	15	US-10-389-640-7	Sequence 7, Appli
38	1263	99.1	323	15	US-10-389-640-13	Sequence 13, Appli
39	1263	99.1	324	14	US-10-314-861-16	Sequence 16, Appli
40	1263	99.1	345	14	US-10-338-411-5	Sequence 5, Appli
41	1263	99.1	345	15	US-10-389-640-5	Sequence 5, Appli
42	1263	99.1	346	14	US-10-338-411-9	Sequence 9, Appli
43	1263	99.1	346	15	US-10-389-640-9	Sequence 9, Appli
44	1263	99.1	359	14	US-10-033-717-33	Sequence 33, Appli
45	1263	99.1	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PPT
; ORGANISM: Aequoria Victoria
US-09-887-784-4

Query Match 100.0%; Score 1274; DB 9; Length 239;
Best Local Similarity 100.0%; Pred. No. 2e-124;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVSKEELFTGVVPIVELDGDVNGHKFSVSGEGEDATYGKLTCLKFCTTCKLPVPWPT	60
Db	1	MVSKEELFTGVVPIVELDGDVNGHKFSVSGEGEDATYGKLTCLKFCTTCKLPVPWPT	60
Qy	61	LVTTLSYGVQCFSRYPDHMKOHDPFKSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTL	120
Db	61	LVTTLSYGVQCFSRYPDHMKOHDPFKSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTL	120
Qy	121	VNRIELKGDIFKEDGNILGHKLEYNHYNHVIYIMADKQNGIKVNFKIRHNIEDSGVQLA	180
Db	121	VNRIELKGDIFKEDGNILGHKLEYNHYNHVIYIMADKQNGIKVNFKIRHNIEDSGVQLA	180
Qy	181	DHYQONTPIGDGPVLLPDNHYLTSQTSALSQDPNEKRDMHVLGFTVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDGPVLLPDNHYLTSQTSALSQDPNEKRDMHVLGFTVTAAGITLGMDELYK	239

This Page Blank (uspto)

RESULT 2

US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; PRIOR FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
IS-10-296-953-4

Query Match 100.0%; Score 1274; DB 12; Length 239;
Best Local Similarity 100.0%; Pred. No. 2e-124;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPWT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGIKDFKEDGNILGHKLEYNYSNHNVIIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRLEKGIKDFKEDGNILGHKLEYNYSNHNVIIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 3

S-10-270-223-6
; Sequence 6, Application US/10270223
; Publication No. US20030143634X1
; GENERAL INFORMATION:
; APPLICANT: BioImage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION RE
; TITLE OF INVENTION: INTERACTIONS B: FLUORESCENCE REDISTRIBUTION.
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/270,223
; PRIOR FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
S-10-270-223-6

Query Match 100.0%; Score 1274; DB 14; Length 363;
Best Local Similarity 100.0%; Pred. No. 3.7e-124;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPWT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGIKDFKEDGNILGHKLEYNYSNHNVIIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRLEKGIKDFKEDGNILGHKLEYNYSNHNVIIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 4

US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match 100.0%; Score 1274; DB 14; Length 893;
Best Local Similarity 100.0%; Pred. No. 1.3e-123;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPWT 60
DB 655 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPWT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRLEKGIKDFKEDGNILGHKLEYNYSNHNVIIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 775 VNRLEKGIKDFKEDGNILGHKLEYNYSNHNVIIMADKQNGIKVNFKIRHNIEDGSVOLA 834
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 893

RESULT 5

US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0

This Page Blank (uspto)

```
/ SEQ ID NO 32
/ LENGTH: 1132
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          100.0%; Score 1274; DB 14; Length 1132;
Best Local Similarity 100.0%; Pred. No. 1.9e-123;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 953

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
/ Sequence 2, Application US/09887784
/ Patent No. US20020177189A1
/ GENERAL INFORMATION:
/ APPLICANT: BJORN, Sara et al
/ TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
/ FILE REFERENCE: 3759-0115P
/ CURRENT APPLICATION NUMBER: US/09/887,784
/ CURRENT FILING DATE: 2001-06-19
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 2
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Aequorea victoria
US-09-887-784-2

Query Match          99.4%; Score 1266; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.4e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
/ Sequence 2, Application US/10296953
/ Publication No. US20040072995A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: BJORN, SARA P.
/ APPLICANT: PAGLIARO, LEN
/ APPLICANT: THASTRUP, OLE
/ TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
/ FILE REFERENCE: PL0095
/ CURRENT APPLICATION NUMBER: US/10/296,953
/ CURRENT FILING DATE: 2002-11-26
/ PRIOR APPLICATION NUMBER: PA 2000 00953
/ PRIOR FILING DATE: 2000-06-19
/ PRIOR APPLICATION NUMBER: 60/212,681
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: 60/290,170
/ PRIOR FILING DATE: 2001-05-10
/ PRIOR APPLICATION NUMBER: PA 2001 00739
/ PRIOR FILING DATE: 2001-05-10
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.4%; Score 1266; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.4e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
/ Sequence 2, Application US/09920922
/ Patent No. US20020083488A1
/ GENERAL INFORMATION:
/ APPLICANT: Miyawaki, Atsushi
/ APPLICANT: Sawano, Asako
/ TITLE OF INVENTION: METHOD FOR MUTAGENESIS
/ FILE REFERENCE: 11283-012001
/ CURRENT APPLICATION NUMBER: US/09/920,922
/ CURRENT FILING DATE: 2001-08-02
/ PRIOR APPLICATION NUMBER: JP 2000-237166
/ PRIOR FILING DATE: 2000-08-04
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          99.1%; Score 1263; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
```

61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKEGDTL 120
61 LVTTLTGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFVTAAGITLGMDELYK 239

RESULT 9
US-09-999-745-4
; Sequence 4, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-999-745-4

Query Match 99.1%; Score 1263; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60
1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60
61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKEGDTL 120
61 LVTTLTGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFVTAAGITLGMDELYK 239

RESULT 10
US-09-866-538-4
; Sequence 4, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT

; ORGANISM: Aequorea victoria
US-09-866-538-4
Query Match 99.1%; Score 1263; DB 10; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60
1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60
61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKEGDTL 120
61 LVTTLTGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFVTAAGITLGMDELYK 239
RESULT 11
US-09-797-496B-2
; Sequence 2, Application US/09797496B
; Publication No. US20030049597A1
; GENERAL INFORMATION:
; APPLICANT: Simon, Sanford M.
; APPLICANT: Chen, Yu
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof
; FILE REFERENCE: 600-1-267
; CURRENT APPLICATION NUMBER: US/09/797,496B
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described
; OTHER INFORMATION: in specification
US-09-797-496B-2

Query Match 99.1%; Score 1263; DB 10; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60
1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60
61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKEGDTL 120
61 LVTTLTGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLVGFVTAAGITLGMDELYK 239
RESULT 12
US-09-794-308-4
; Sequence 4, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA

APPLICANT: TSIEH, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REG1530
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match 99.1%; Score 1263; DB 10; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFIRHNIEDGSVQLA 180
QY 181 DHYQNTPIGDPVLLPDNHYLSTQSALSQSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQNTPIGDPVLLPDNHYLSTQSALSQSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
Sequence 4, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEH, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIO-METRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REG1550
CURRENT APPLICATION NUMBER: US/09/865,291
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match 99.1%; Score 1263; DB 10; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFIRHNIEDGSVQLA 180

QY 181 DHYQNTPIGDPVLLPDNHYLSTQSALSQSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQNTPIGDPVLLPDNHYLSTQSALSQSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US20030212265A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Acsushi
APPLICANT: Ilopis, Juan
APPLICANT: Wachter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR APPLICATION NUMBER: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 99.1%; Score 1263; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFIRHNIEDGSVQLA 180
QY 181 DHYQNTPIGDPVLLPDNHYLSTQSALSQSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQNTPIGDPVLLPDNHYLSTQSALSQSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
RESULT 15
US-10-121-258-13
Sequence 13, Application US/10121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP1
CURRENT APPLICATION NUMBER: US/10/121,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24

; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match 99.1%; Score 1263; DB 14; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MVSKEELFTGVDPILVELDGDVNGHKFVSVEGEGDATYCKLTKEICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVDPILVELDGDVNGHKFVSVEGEGDATYCKLTKEICTTGKLPVPWPT 60
Qy 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:28
Job time : 35.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds
(without alignments)
2224.817 Million cell updates/sec

Title: US-09-887-784-222L
Perfect score: 1274
Sequence: 1 MVSKGEELFTGVVPILVELD.....VLIGFTVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PTR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1241	97.4	238	1 JQ1514	green-fluorescent
2	1105	8.2	785	2 H72228	hypothetical prote
3	91.5	7.2	861	2 H64102	leucine-trna ligase
4	90	7.1	632	2 T06586	DNA-binding protei
5	89.5	7.0	887	2 E82590	leucyl-trna synthet
6	88.5	6.9	655	2 D83917	DNA topoisomerase
7	87.5	6.9	370	2 E70390	iron-sulfur cofact
8	87.5	6.9	860	2 AC0582	leucyl-trna synthet
9	87.5	6.9	2252	1 A36028	DNA-directed DNA p
10	87.5	6.9	2573	2 D71614	hypothetical prote
11	87	6.8	578	1 I40794	dihydroliopamide d
12	87	6.8	874	2 JC4930	S-layer protein pr
13	86.5	6.8	797	2 JC4078	protective surface
14	86.5	6.8	808	2 F64102	protective surface
15	86	6.8	357	2 G81355	tRNA (uracil-5)-m
16	85.5	6.7	788	1 JDUVLH	DNA-directed DNA p
17	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
18	85.5	6.7	941	2 S29043	cellulase (EC 3.2.
19	85	6.7	281	2 AD2052	hypothetical prote
20	85	6.7	439	2 JH0414	synaptogamin o-p65
21	84.5	6.6	425	2 C97354	hypothetical prote
22	84.5	6.6	613	2 A99552	oligohydroxyprote
23	84	6.6	353	2 E84941	imidazoleglycerol-
24	83.5	6.6	836	1 JDUVLD	DNA-directed DNA p
25	83.5	6.6	1134	2 A60234	IGA FC receptor pr
26	83.5	6.6	1164	1 FCSOAG	IGA FC receptor pr
27	83	6.5	461	2 T27856	photosystem II chl
28	83	6.5	471	2 T27856	hypothetical prote
29	83	6.5	774	2 T39539	alpha-amylase homo

ALIGNMENTS

RESULT 1

JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:1347277
A:Accession: JQ0692
A:Molecule type: DNA
A:Residues: 1-107, 'S', 109-238 <PRA1>
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99, 'P', 'F', '101-140, 'L', '142-218, 'V', 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>
R:Inouye, S.; Teuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 'Q', '26-156, 'P', 158-171, 'K', 173-238 <INO>
A:Cross-references: GB:L29145; NID:G606383; PIDN:AAA58246.1; PID:G606384
R:Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13, 'V', 15-24, 'O', '26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009
A:Experimental source: clone gfp1
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24, 'Q', '26-29, 'R', '31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 209-238 <PRA5>
A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011
A:Experimental source: clone gfp2
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65692; PDB:IGFL
A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-91
A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58953; MUID:98294543; PMID:9631087

A;Contents: annotation; X-ray crystallography, 1.9 angstroms
C;Comment: This protein is excited by the photoprotein seagurin (see PIR:AQJFNV) emitting
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-tyr-
C;Genetics:
A;Gene: GPP
A;Introns: 69/3; 167/3
C;Superfamily: green-fluorescent protein
C;Keywords: Chromoprotein; luminescence
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.4%; Score 1241; DB 1; Length 238;
Best Local Similarity 97.1%; Pred. No. 1.2e-36;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVVELDGDVNGHKFSVSGEGGDATYKGLTKFKICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPIVVELDGDVNGHKFSVSGEGGDATYKGLTKFKICTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELAGIDPKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELAGIDPKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLGFTVTAAGITLGMDELYK 238

RESULT 2
H72228
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: H72228
A;Title: Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: H72228
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-785 <ARN>
A;Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1624

Query Match 8.2%; Score 105; DB 2; Length 785;
Best Local Similarity 19.7%; Pred. No. 0.88;
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPIVVELDGDVNGHKFSVSGEGGDATYKGLTKFKICTTGKLPVWPPTLV 62
DB 15 NEGRFSPEGTVPVGVQAD-----LVRKGLLPHVYVGM- 46

QY 63 TTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLVN 122
DB 47 -----NEDLFEIEDRWIYVEREFKEDVKEGERVDLVEGVDTLN 88

QY 123 RIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFIRHNIEDGSVOLADH 182
DB 89 DVYLVGYL---GSTEDMFIEYRFDVTNLV---KEKNHLKVYIK-----SPIRVPKT 134

QY 183 YQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLGFTVTAAGITLGM 235
DB 135 LEQNYGVLGPF-----BDP-----IRGYRKQAQSYGWD 163

RESULT 3

H64102
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)

A;Alternate names: leucyl-tRNA synthetase

C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002

C;Accession: H64102

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: H64102

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-861 <TIGR>

A;Cross-references: GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC22581.1; PID:g1573943; T;

C;Genetics:

A;Gene: leuS

C;Superfamily: leucine-tRNA ligase

C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.2%; Score 91.5; DB 2; Length 861;
Best Local Similarity 24.1%; Pred. No. 13;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVWPPTLVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFK 103
DB 314 TGDKLPI-WVANFVLMHGYTGAVMAVPAH-DQDRD-----EFAQKYSLPINQVIAPLA 364

QY 104 DGNKYTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNVSHNVIMADK-OKNGI 162
DB 365 DEEIDLTKQAFVEHGHKLVNSDFDGKNF--DGAFF-----IADKLEKLG 408

QY 163 ---KVNFKIRH-----NIEDGSVOLADHYQQNTPIGDGPVLLPDNHYL- 202
DB 409 GKRQVNYRLRDWGVSRQRYWGAPIPMLTLENGDVVPA-----PWEDLPILLPEDVWMD 461

QY 203 STQSALSKDPN 213
DB 462 GVKSPINADPN 472

RESULT 4

T06586

DNA-binding protein PD2 - garden pea

C;Species: Pisum sativum (garden pea)

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999

C;Accession: T06586

R;Sato, N.; Kazuno, A.A.; Ohta, N.; Ohshima, K.

submitted to the EMBL Data Library, June 1996

A;Description: Identification of a novel family of DNA-binding proteins with two AT-hook

A;Reference number: Z15774

A;Accession: T06586

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-632 <SAT>

A;Cross-references: EMBL:X98740; NID:e995229; PIDN:CAA67292.1; PID:e275185

A;Experimental source: cv. Alaska

Query Match 7.1%; Score 90; DB 2; Length 632;
Best Local Similarity 23.3%; Pred. No. 12;
Matches 49; Conservative 26; Mismatches 79; Indels 56; Gaps 7;

QY 16 LVLEDGDVNGHKFSVSGEGGDATYKGLTKLTKFKICTTGKLPVWPPTLVTTLSYGVCFSRY 75
DB 363 IVQGRDVGSKVDVINKESNEATIPENK----PTBPKLDVEQELAAATT----- 408

QY 76 PDHMKQHDFFKSAMPEGYV-----QERTIFFKDDGNKYKTRAEVKFEGDTLVNRIE 125

Db 409 PPSAKNVNLT KD L I V T F P L R S V A R T S G R E G S E E L K D S N S L E R D T K K L E L E Q G K N S - E 467
QY 126 LKIGDFKEDGNILGHKLE-----NYNSHNYYIMADKQK-----N 160
Db 468 LKGIETDNTSLLDKFEFNAIKNKILKEISNPHDVSANHSHTNKQVTVSHQKAEITNN 527
QY 161 GIKVNFKIRHNIEDG-----SVOLADHYQ 184
Db 528 QSQVEDVAKNKIQDQSKPSESLHKADKYR 557
RESULT 5
E82590
leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: E82590
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82590
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-887 <SIM>
A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.B.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, R.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2176
C:Superfamily: leucine-tRNA ligase
Query Match 7.0%; Score 89.5; DB 2; Length 887;
Best Local Similarity 22.2%; Pred. No. 21;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;
QY 50 TTGKLPVPWPTLVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107
Db 329 TNEQLPV-WVANFVLMAYGTGAVMVGPHDQDQEF--ANKYGLPIRQVIALKEPKNQDE 385
QY 108 -----KTRAEVKFPGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVI 153
Db 386 STWEPDVRDWDYADKTR---EFE---LINSAEFDGLDYQDAFVLAERFE----- 429
QY 154 MADKQKNG-IKVNFKIRHNIEDGSVOLADHYQOQNTPI-----GGGPVLLPDN 199
Db 430 ---RQGRQRVNNYLR-----DWGVSQRVWGCPIPVIYCTCGAVPVPEDQLPVILPEN 482
QY 200 -HYLSLTQSALSCKDPNKR 216
Db 483 VAFSGTGSPIKTPDPEWRK 500
RESULT 6
D83917
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: D83917

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: D83917
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <STO>
A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05859.1; GSPDB:GN001
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2140
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B
Query Match 6.9%; Score 88.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;
QY 22 DVNGHK---FSVSGEGEGDAT---YGLTKLTKPI-----CTTGKLPVPWP 59
Db 63 NVTIHKQSVSVRDEGRGMPTGMHKLKPTPEVILTVLHAGGKFGQGVATSGLHGVA 122
QY 60 TLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQER-----TIFKDDG----- 105
Db 123 SYVNALSEWLIVEIKRDGWWYEQRFENGKPKSTTLEKKGKTRQGTGTHFKPDTVFSST 182
QY 106 --NYKTRAEVKFPGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADK----- 157
Db 183 NFNVETLSRLREAAFLKGLKIELVDLDDITKEVFH-YEDGKAFVEYLNEDEKTLHPV 241
QY 158 -----QKNGIKVNFKIRHNIEDGSVOLADHYQOQNTPIGDPVLLPDNHYLSLTQSALS 212
Db 242 VFFNGESNGIEIEFAFOFN--DGYTENVLFSVNNVVRTKD-----GTHELGAKTAMTRAV 294
QY 213 NE 214
Db 295 NE 296
RESULT 7
E70390
iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus
N:Contains: L-cysteine sulfurtransferase (EC 2.8.1.-)
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999
C:Accession: E70390
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: E70390
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-370 <AQF>
A:Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: nifs1
C:Superfamily: nitrogen fixation protein nifs
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase
F:195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted
F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted
Query Match 6.9%; Score 87.5; DB 2; Length 370;
Best Local Similarity 25.4%; Pred. No. 9.7;
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;
QY 4 KGEELFTGVV----PILVELD----GDVNGHKF-SVSGEG-----EGDATYKLTAKFICT 50
Db 164 KGVPVLTDAVQAIGKPIELKNSIYATFSGHKPHAIKSGSGLYISDEANYEPLVGGQGE 223
QY 51 TGKLP-----VWPVTLVTTLSYGVQCFSRYPDHMKQ-HDFFKSAMPEGYVQERTIFFKOD 104

A;Residues: 1-874 <ZHU>
A;Cross-references: GB:U38842; NID:g1055336; PIDN:AAC44405.1; PID:g1055337
A;Experimental source: strain NM105
C;Comment: This protein is a glycoprotein. It functions as protective coats, molecular s
C;Genetics:
A;Gene: olpA
C;Superfamily: S-layer repeat homology
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-874/Product: S-layer protein #status predicted <MAT>
F;157-209/Domain: S-layer repeat homology <SLR3>

Query Match 6.8%; Score 87; DB 2; Length 874;
Best Local Similarity 22.8%; Pred. No. 33;
Matches 65; Conservative 28; Mismatches 96; Indels 96; Gaps 14;

Qy 6 EELF--TGWPIL---VELDGDVNGHKFSVSGEGEDAT-----YGLTLTKFI 48
Db :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: |||||
551 KEVPQTGVVKVLDTVTNEGSIGTSSIKVGKENVGAGITHFNQPNASGEGYGSLSHVEVT 610

Qy 49 CTTGKLPPWPTLV-----TTLSYGVCQFSRYPDHMKQHDFPKSAMPGYVQER 97
Db :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: |||||
611 KSNIGHEAPRELVSQAKGGEADTTLGAGNTVAQLSNYYTEGVYADAADLAGY--EF 668

Qy 98 TIFPKDDGNKYTRAEVKEFGDTLVNRIELKGID---FKEDGNILGHKLEYNNSHNVIYM 154
Db :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: |||||
669 RV-----GNDKI-ASAIEGTGLKVGTAGTVTDILTCKDGATAGH-ATTIVTOENIQIT 721

Qy 155 ADKOK-----NGIKNVFKIRHN---IEDGSVQLADH 182
Db :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: |||||
722 SVKFQDEVEGFENRKVNIDRVLDVVKSDDVLNGIKLNISTEHKVRIVDEGTEQ---- 777

Qy 183 YQNTPPIGDGPVLLPDNNHYLSTQSALSQDPNEKRDMVLLGFVTA 227
Db :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: |||||
778 -----GKV-----YLDRDNATFDGND-----VALGYVTA 802

RESULT 13
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C;Species: Haemophilus influenzae
A;Variety: type b
C;Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
R;Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A;Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl
A;Reference number: JC4078; MUID:95255676; PMID:7737523
A;Accession: JC4078
A;Molecule type: DNA
A;Residues: 1-797 <FLA>
A;Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448
A;Experimental source: type b
C;Superfamily: protective surface antigen D-15
C;Keywords: surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 6.8%; Score 86.5; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. No. 32;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

Qy 65 LSYGVCQFSRYPDHMKQHDF-----FKSAMPEGYVQE-----RTI 99
Db :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: |||||
427 IGYTESGISYOASVKQDNFLGTGAAVSTAGTKNDYGTSVNLGYTEPYFTKDGVSLGGNV 486

Qy 100 FFKDDGNKYTRAEVKEFGDTLVNRIELKGIDPKEDGNI---LGH-----KLEYNYS 148
Db :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: |||||
487 FFENYDNSKDSSTSSNYKRTTYGTSNVTL-GFPVNNNSYVVLGHTYKNKISNFALAYN--- 542

Qy 149 HNYVMADKQK-NGIKNVFKIRHNIEDGSVOLADHYQQ-----NTPIDGPGVLL 196
Db :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: ||||| :|: |||||
543 RNLYIQSMKFGNGIKTN-----DFDFSFGWNYSNLRGYFFPTKGVKASLG-GRVTI 593

Qy 197 P--DNHYLSQTQSALSQDPNEKRDMHVLGFTVTAAGITLG 233
| | | | : : : : |
| | | | : : : : |
Db 594 PGSDNKYYKLSADVQGFYPLDRDHLWVWSAKASAGYANG 632

RESULT 14
F64102
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C/Accession: F64102
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
Gocayne, J.B.; Scott, J.; Shirley, R.; Liu, L.I.; Glodsk, A.; Kelley, J.M.; Weidman, J.;
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: F64102
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-808 <TIGR>
A/Cross-references: GB:L42023; TIGR:HI0917
C/Superfamily: protective surface antigen D-15
C/Keywords: surface antigen

	Query Match	6.8%; Score 86.5; DB 2; Length 808;
	Best Local Similarity Matches 48; Conservative	21.9%; Pred.No.33; 78; Indels 63; Gaps 11;
Qy	65 LSYGVQCFSRYPDHMKQHDF-----FKSAMPEGYVQE-----RTI 99	
Dd	440 IGYTESGISVQSARKDNFLGTCAAVSIAGTKNDYGTSVLNGTEPFYFTKDGVSGLGNV 499	
Qy	100 PFKDDGNYKTRAEVKFEGDITLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148	
Dd	500 PFENYDKSKDSITSNYKRITYTGSNVTL-GFPVNENNYYVGLTHTYKNKSINFALEYN- 555	
Qy	149 HNVYTMDKK-QGIKVNFKIRHNEIGSVQLADHYQQ-----NTPIGDGPVLL 196	
Dd	556 RNLYIQSMKFCKGIIKTN-----DPDFSGWNYSLNRGFYPTKGVKASLG-GRVTI 606	
Qy	197 P--DNHYLSTOALS-KDPNEKRDMHVLLGGFVTAAGITILG 233	
Dd	607 PGSDNNKYIKLAGDVQGFGFPLDRDLHLVVVSAKASAGYANG 645	

RESULT 15

G81355
tRNA (uracil-5)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter jejuni
S:Species: Campylobacter jejuni
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: G81355
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, S.; Whitehead, P.; MacArthur, A.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell, C.W.; Quail, M.; Parkhill, J.
Nucleotide 403, 665-668, 2000
A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence
A:Reference number: AB1250; PMID:20150912; PMID:10688204
A:Accession: G81355
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <PAR>
A:Cross-references: GB:AL111168; CB:AL111168; NID:G96968128; PIDN:CAB73096.1; PID:G9696827
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: trmA; Cj0831c
K:Keywords: methyltransferase; S-adenosylmethionine

Query Match 6.8%; Score 86; DB 2; Length 357;
 Best Local Similarity 24.8%; Pred. NO. 12;
 Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;
 QV R0 KOHDFEKGAMPGVYVOERTIFFKDDGNKYTRAEVKF--EGDTLV-----NRRLKG 128

Db	14	EKHSFIKKYFKFYTKDFKLFASDKDCHYTRRAELUSFYHENDTLFAMFDPKSKKXVILEY	73
Qy	129	IDFKED-----GNILGHKLEYNYSNHYIWMADKONGIKVNFKIRHNIE	173
Db	74	LDFADEKICAPMRLLFVLRQDNKLKEKL-----FGVEFLTTKQE--LSITLLYHKNIE	125
Qy	174	D 174	
Db	126	D 126	

Search completed: June 21, 2004, 16:01:57
Job time : 10.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-2221
Perfect score: 1274
Sequence: 1 MYSKGELFTGVVPILVELD.....VLGLFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1247	97.9	238	1	GFP_AEQVI
2	91.5	7.2	861	1	SYL_HAEIN
3	89.5	7.0	879	1	SYL_XYLFA
4	87.5	6.9	860	1	SYL_SALTI
5	87.5	6.9	860	1	SYL_SALTY
6	87.5	6.9	2222	1	DPOE_YEAST
7	87	6.8	689	1	AC2L_HUMAN
8	87	6.8	874	1	SLAP_BACLI
9	86.5	6.8	533	1	CP51_CANGA
10	86.5	6.8	795	1	D152_HAEIN
11	86.5	6.8	797	1	D151_HAEIN
12	86.5	6.8	879	1	SYL_XYLFT
13	86.5	6.8	1603	1	VIT4_CABEL
14	86	6.8	357	1	TRMA_CAMJE
15	85.5	6.7	788	1	DPOL_HPBHE
16	85.5	6.7	793	1	D153_HAEIN
17	85.5	6.7	886	1	ITH3_MSAU
18	85.5	6.7	941	1	GUN_BACS6
19	85	6.7	439	1	SY62_DISOM
20	84.5	6.6	501	1	AMPA_WICBR
21	84.5	6.6	613	1	PEPF_WYCPU
22	84.5	6.6	859	1	SYL_SHEON
23	84	6.6	353	1	HIS7_BUCAI
24	84	6.6	366	1	SET7_HUMAN
25	84	6.6	504	1	YC03_KLEPN
26	83.5	6.6	538	1	GRBE_RAT
27	83.5	6.6	658	1	ADAS_HUMAN
28	83.5	6.6	1164	1	BAG_STRAG
29	83	6.5	461	1	PSBC_CTRAP
30	83	6.5	774	1	AMF2_SCHPO
31	82.5	6.5	533	1	NIFD_CLOPA
32	82	6.4	682	1	PRC_ECOLI
33	82	6.4	752	1	NECI_RAT

34	81.5	6.4	589	1	SYD_HAEDU
35	81	6.4	336	1	YD48_METJA
36	81	6.4	682	1	AC2L_MOUSE
37	81	6.4	1224	1	COPA_HUMAN
38	80.5	6.3	393	1	TRME_HELPY
39	80.5	6.3	658	1	ADAS_CAVPO
40	80.5	6.3	860	1	SYL_ECOS7
41	80.5	6.3	860	1	SYL_ECOL6
42	80.5	6.3	860	1	SYL_ECOLI
43	80	6.3	461	1	PSBC_CHLEU
44	80	6.3	737	1	OPT1_DROME
45	79.5	6.2	312	1	TRXB_CHDMU

ALIGNMENTS

RESULT 1					
GFP_AEQVI					
ID	GFP_AEQVI	STANDARD;	PRT;	238 AA.	
AC	P42212; Q17104; Q27903;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Green fluorescent protein.				
GN	GFP.				
OS	Aequorea victoria (Jellyfish).				
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;				
OC	Aequoreidae; Aequorea.				
OX	NCBI_TaxID=6100;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RX	MEDLINE=92175527; PubMed=1347277;				
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,				
RA	Cormier M.J.;				
RT	"Primary structure of the Aequorea victoria green-fluorescent				
RT	protein."				
RL	Gene 111:229-233(1992).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=94185810; PubMed=8137953;				
RA	Inouye S., Tsuji F.I.;				
RT	"Aequorea green fluorescent protein. Expression of the gene and				
RT	fluorescence characteristics of the recombinant protein.";				
RL	FEBS Lett. 341:277-280(1994).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97299832; PubMed=9154981;				
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;				
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent				
RT	protein by modification of its codon usage.";				
RL	Plant Mol. Biol. 33:989-999(1997).				
RN	[4]				
RP	CHROMOPHORE.				
RX	MEDLINE=93192221; PubMed=8448132;				
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;				
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea				
RT	green-fluorescent protein.";				
RL	Biochemistry 32:1212-1218(1993).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).				
RX	MEDLINE=96355665; PubMed=8703075;				
RA	Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,				
RA	Remington S.J.;				
RT	"Crystal structure of the Aequorea victoria green fluorescent				
RT	protein."				
RL	Science 273:1392-1395(1996).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).				
RX	MEDLINE=98294543; PubMed=9631087;				
RA	Yang F., Moss L.G., Phillips G.N. Jr.;				
RT	"The molecular structure of green fluorescent protein.";				
RL	Nat. Biotechnol. 14:1246-1251(1996).				

Q7vnf0	haemophilus
Q58743	methanococc
Q99nb1	mus musculus
P53621	homo sapien
O25443	helicobacte
P97275	cavia porce
Q8xbn8	escherichia
Q8fjy9	escherichia
P07813	escherichia
Q08684	chlamydomon
P91679	drosophila
Q9pck7	chlamydia m

[7]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
 RX MEDLINE=98455509; PubMed=9782051;
 RA Wachter R.M., Eisliger M.A., Kallio K., Hanson G.T., Remington S.J.;
 RT "Structural basis of spectral shifts in the yellow-emission variants
 of green fluorescent protein.";
 RL Structure 6:1267-1277(1998).
 [8]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=99238303; PubMed=10220315;
 RA Eisliger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
 RT "Structural and spectral response of green fluorescent protein
 variants to changes in pH";
 RL Biochemistry 38:5296-5301(1999).
 CC -1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the
 blue chemiluminescence of the protein aequorin into green
 fluorescent light by energy transfer. Fluoresces in vivo upon
 receiving energy from the Ca(2+)-activated photoprotein aequorin.
 CC Absorbs light maximally at 395 nm and exhibits a smaller
 absorbance peak at 470 nm. The fluorescence emission spectrum
 peaks at 509 nm with a shoulder at 540 nm.
 CC -1- SUBUNIT: Monomer.
 CC -1- TISSUE SPECIFICITY: Photocytes.
 CC -1- PTM: Contains a covalently attached chromophore, which is composed
 of modified amino acid residues. The chromophore is formed upon
 cyclization of the residues Ser-dehydroTyr-Gly.
 CC -1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
 chimeric proteins of GFP linked to other proteins where it
 functions as a fluorescent protein tag. GFP tolerates N- and C-
 terminal fusion to a broad variety of proteins. It has been
 expressed in bacteria, yeast, slime mold, plants, Drosophila,
 zebrafish, and in mammalian cells. As a noninvasive fluorescent
 marker in living cells, it allows for a wide range of applications
 where it may function as a cell lineage tracer, reporter of gene
 expression, or as a measure of protein-protein interactions.
 CC -1- DATABASE: NAME-Protein Spotlight;
 CC NOTE=Issue 11 of June 2001;
 CC WWW="http://www.expasy.org/spotlight/articles/sptl011.html".

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; M62654; AAA27722.1; -;
 DR EMBL; M62653; AAA27721.1; -;
 DR EMBL; L29345; AAA58246.1; -;
 DR EMBL; X96418; CAA65278.1; -;
 DR FIR; JS0692; JQ1514.
 DR PDB; 1B9C; 17-NOV-00.
 DR PDB; 1BFP; 07-JUL-97.
 DR PDB; 1C4F; 14-JUN-00.
 DR PDB; 1EWA; 08-NOV-96.
 DR PDB; 1EMB; 16-JUN-97.
 DR PDB; 1EMC; 20-AUG-97.
 DR PDB; 1EME; 20-AUG-97.
 DR PDB; 1EMF; 20-AUG-97.
 DR PDB; 1EMG; 12-MAY-99.
 DR PDB; 1EMK; 20-AUG-97.
 DR PDB; 1EML; 20-AUG-97.
 DR PDB; 1EMW; 20-AUG-97.
 DR PDB; 1F09; 17-NOV-00.
 DR PDB; 1F0B; 17-NOV-00.
 DR PDB; 1GFL; 11-JAN-97.
 DR PDB; 1HGT; 15-JAN-02.
 DR PDB; 1HUY; 04-JUL-01.
 DR PDB; 1JBY; 07-JAN-03.
 DR PDB; 1JBJ; 28-AUG-02.
 DR PDB; 1KP5; 28-AUG-02.
 DR PDB; 1KYP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.
 DR PDB; 1KYS; 10-APR-02.
 DR PDB; 1YFP; 28-OCT-98.
 DR PDB; 2EMD; 20-AUG-97.
 DR PDB; 2EMN; 20-AUG-97.
 DR PDB; 2EMO; 20-AUG-97.
 DR InterPro; IPR000786; GFP like.
 DR InterPro; IPR009017; GFP like.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFLUORESCENT.
 DR ProDom; PD013756; Green_fl_protein; 1.
 KW Luminescence; 3D-structure.
 FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).
 FT MOD_RES 66 66 2,3-DIDEHYDROTYROSINE.
 FT VARIANT 100 100 F -> Y.
 FT VARIANT 108 108 T -> S.
 FT VARIANT 141 141 L -> M.
 FT VARIANT 219 219 V -> I.
 FT CONFLICT 2 2 S -> G (IN REF. 3).
 FT CONFLICT 25 25 H -> Q (IN REF. 2).
 FT CONFLICT 80 80 Q -> R (IN REF. 3).
 FT CONFLICT 157 157 Q -> P (IN REF. 2).
 FT CONFLICT 172 172 E -> K (IN REF. 2).
 FT HELIX 4 8
 FT STRAND 12 22
 FT TURN 23 24
 FT STRAND 25 36
 FT TURN 37 40
 FT STRAND 41 48
 FT TURN 49 50
 FT HELIX 57 60
 FT TURN 61 63
 FT HELIX 69 71
 FT STRAND 73 73
 FT HELIX 76 81
 FT HELIX 83 86
 FT TURN 87 90
 FT STRAND 92 100
 FT TURN 101 102
 FT STRAND 105 115
 FT TURN 116 117
 FT STRAND 118 128
 FT TURN 132 133
 FT TURN 135 139
 FT STRAND 141 141
 FT STRAND 148 155
 FT TURN 156 159
 FT STRAND 160 171
 FT TURN 172 173
 FT STRAND 176 187
 FT STRAND 199 208
 FT TURN 211 212
 FT STRAND 217 227
 SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21BFB6E05 CRC64;
 Query Match 97.9%; Score 1247; DB 1; Length 238;
 Best Local Similarity 98.3%; Pred. No. 4.7e-96;
 Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKFKICTTGKLPVPWPTL 61
 :|||||
 Db 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKFKICTTGKLPVPWPTL 60
 QY 62 VTTLISYGVQCFRYPDHMKQHPFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
 :|||||
 Db 61 VTTFISYGVQCFRYPDHMKQHPFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
 QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHVYIMADKQKGIKNFKIRHNIEDGSVQLAD 181
 :|||||
 Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHVYIMADKQKGIKNFKIRHNIEDGSVQLAD 180
 QY 182 HYQONTPIGDPVLLPDNHYLTQSALS KDPNEKRDMVLLGFVTAAGITLGMDELYK 239
 :|||||

Db 181 HQQNTPIGDPVLPDHNHLSYQSALSKPDNKRDKHNVLLFEVTAAGITGHMDLYK 238

RESULT 2
SYL_HAEN ID SYL_HAEN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeURS).
GN LEUS OR HI0921
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
ON NCBI_TaxID=727;
RX STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "whole-genome random sequencing and assembly of Haemophilus influenzae Rd";
RT Science 269:496-512(1995).
RL -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP + diphosphate + L-leucyl-tRNA(Leu).
CC CC
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
DR EMBL: U32774; AAC22581.1; --
DR PIR: H64102; H64102.
DR TIGR: HI0921; --
DR HAMAP: MF 00049; 1.
DR InterPro: IPR002302; Leu-trNASyntia.
DR InterPro: IPR002300; tRNA-synt_1a.
DR InterPro: IPR001412; tRNA-synt_1.
DR InterPro: IPR009008; ValRS_fleRS_edit.
DR Pfam: PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMS; TIGR00396; leus bact. 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT BINDING 619 623 "KMSX" REGION.
FT BINDING 622 623 ATP (BY SIMILARITY).
SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 6.5;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
50 TTGKLVPWPPTLVTLISYGVCFSRYDPDMKHQDFKSNMPEGYVOERTIFFKD----- 103
314 TGDKLPIT-WVANFVLHMHYGTGVAMVPAHF-DORDF-----EFAQKYSLPKIQVIAPLA 364

Qy 104 DGNKYTRAIVEKEGDTLVNRTELKGIDFKEDGNILGHKLEYNNSHNYINADK-QKNGI 163
Db 365 DEEDILTQAFVEHGKLVNSDEFDGKNF--DCAFNG-----TADKLEKLGV 408
Qy 163 ---KVNFKIRH-----NIEDGSVOLADHYQONTPIGDPVLLPNHYL- 202
Db 409 GKRVNRYLRLDWGSVRQRWGAPIPMLTLENDVVPA-----PMEDLPILLPEDDVMD 461
Qy 203 STQSALSCKDPN 213
Db 462 GVKSPINADPN 472
RESULT 3
SYL_XYLEFA ID SYL_XYLEFA STANDARD; PRT; 879 AA.
AC Q9PBGB;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeURS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.I., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Francis S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieser J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.A.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesqueira J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.P., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.W., Tshuko M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa";
Nature 406:151-159(2000).
RL -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP + diphosphate + L-leucyl-tRNA(Leu).
CC CC
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC EMBL: AF004031; AAF84975.1; ALT_INIT.
DR HAMAP: MF 00049; 1.


```

RA RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamiyama K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Irie R., Sato H.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Murakawa K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Ota T., Hayashi K., Sugiyama T., Otsuki T., Ishibashi T.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S.,
RA Kawai Y., Wakamatsu A., Kanehori K., Suzuki Y., Sugano S.,
RA Nagahari K., Masuho Y., Nagai K., Isogai T.;
RA "NEBO human cDNA sequencing project.";
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RL
RN [4]
RN SEQUENCE OF 336-689 FROM N.A. (ISOFORM 1).
RP
RP TISSUE=Brain;
RC MEDLINE=21245130; PubMed=11347906;
RX
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
CC
CC -!- FUNCTION: Converts acetate to acetyl-CoA so that it can be used
CC for oxidation through the tricarboxylic cycle to produce ATP and
CC Co(2) (By similarity).
CC
CC -!- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
CC acetyl-CoA.
CC
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing, Named isoforms=2;
CC
CC Name=1;
CC
CC Name=2;
CC
CC IsoID=Q9NUB1-1; Sequence=Displayed;
CC
CC IsoID=Q9NUB1-2; Sequence=VSP 007249;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC
CC -!- CAUTION: Ref.1 (CAB81884) sequence differs from that shown due to
CC erroneous gene model prediction.
CC
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
CC frameshift in position 250 and numerous sequencing errors.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; AL035661; CAB75500.1; -
CC EMBL; AL080312; CAB81884.1; ALT_SEQ.
CC EMBL; BC039261; AAH39261.1; -
CC EMBL; BC044588; AAH44588.1; -
CC EMBL; AK027817; BAB55390.1; ALT_INIT.
CC EMBL; AK092295; BAC03853.1; ALT_SEQ.
CC EMBL; AB058749; BAB4475.1; -
CC Genew; HGNC:16091; ACAS2L.
CC InterPro; IPR000873; AMP-bind.
CC Pfam; PF00501; AMP-binding; 1.
CC PROSITE; PS00455; AMP_BINDING; 1.
CC Ligase; Mitochondrial; Transit peptide; Alternative splicing.
FT TRANSIT 1 36 MITOCHONDRION (POTENTIAL)
FT CHAIN 37 689 ACETYL-COENZYME A SYNTHETASE 2-LIKE.
FT DOMAIN 45 53 POLY-ALA.
FT VARSPLOC 446 447 Missing (in isoform 2).
FT
FT CONFLICT 277 277 /FTid=VSP 007249.
FT CONFLICT 488 488 V -> M (IN REF. 2; AAH39261).
FT CONFLICT 488 488 V -> M (IN REF. 2; AAH44588).
SQ SEQUENCE 689 AA; 74856 MW; 66E84E39302AD08B CRC64;

Query Match 6.8%; Score 87; DB 1; Length 689;
Best Local Similarity 24.1%; Pred. NO. 12;
Matches 33; Conservative 16; Mismatches 36; Indels 52; Gaps 7;

```

```
QY 9 FTGVVPIVLVDGNGHGFVSGBEGDATYGGKLTLCFKITCTTGKLPVPMPTLVTLTSYG 68
Db 473 FFGIVPVLMDKGSV-----VEGNSVSGALCIS-----QAWPGWARTI--- 510
QY 69 VQCFSEYPPDHMKQHDFFKSAMPQGYVQERTIFFKDDGNKYKTRA---EVKFEGETLVNRIE 125
Db 511 -----YGDHQRFVDAYFRAYP-GY-----YFTGDGAVRTEGGYQITGRMDVDVI----- 553
QY 126 LKGIIDFKEDGNILGHKL 142
Db 554 -----NISGHL 560

RESULT 8
SLAP_BACLI STANDARD; PRT; 874 AA.
AC P49052; 1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE S-layer protein precursor (Surface layer protein).
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NM 105;
RX MEDLINE=97082965; PubMed=8964497;
RA Zhu X., McVeigh R.R., Malathi P., Ghosh B.K.;
RT "The complete nucleotide sequence of the Bacillus licheniformis NM105
RT S-layer-encoding gene.";
RL Gene 173:189-194 (1996).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38842; AAC44405.1; -.
DR F01; JC4930; JC4930.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
KW Signal; Cell wall; S-layer; Repeat.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 874 S-LAYER PROTEIN.
FT DOMAIN 33 93 SLH 1.
FT DOMAIN 94 155 SLH 2.
FT DOMAIN 156 217 SLH 3.
SQ SEQUENCE 874 AA; 92734 MW; EFADCAAF27D32AF CRC64;

Query Match 6.8%; Score 87; DB 1; Length 874;
Best Local Similarity 22.8%; Pred. No. 16;
Matches 65; Conservative 28; Mismatches 96; Indels 96; Gaps 14;

QY 6 BELF--TGVPVPII---VELDGDVNGHGFVSGBEGDAT-----YCKLTKEI 48
Db 551 KEVFPQTGVKVLVDVTTTNEGSGTSSIRKGVNGVAGTIHFQNPASGEGYGSLSHVEVT 610
QY 49 CTITGKLPVPMPTLV-----TTLSYGVCFSRYPDHMKQHDFFKSAMPQGYVQER 97
Db 611 KSNIGHEAPRLVSVKAGQGEADTILGNNTVAQLSNYTTGEGYVADADLAGY--EF 668
QY 98 TIFFKDDGNKYKTRAEVFKFEGDITLVNRIELKGD---FKEDGNILGHKLEYNVSHNYVM 154
Db 669 RV-----GNDKI-ASAKIEGKTLKVTGKTAGVTDVILTKDGATAGH-ATITVTQENIQIT 721
```

```
QY 155 ADQK-----NGIKVNFKIRHN-----IEDGVSQVLADH 182
Db 722 SVKFDVEVEQENRKVNIDRVLDVVKSKDDVINGIKLNISTEHKVRIVDEGTEQ----- 777
QY 183 YQONTPIGDPVLLPNDHLYSTQSALSQKDPNKRDRHMLVGLFVTA 227
Db 778 -----GKV-----YLDNRDNATFDGND-----VALGYVTA 802

RESULT 9
CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPL1) (P450-LIA1) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717 (1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (L1A1) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907 (1994).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesterol-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-9,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L40389; AAB02329.1; -.
DR EMBL; S75389; AAB32679.1; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
FT Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
```

```
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6BF7 CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.4%; Pred. No. 9.5;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;
QY 25 GHKFSVS--GEGEGDATYKGLTKFKICTTGKLPVPEWPTLVTLTSLYGVQCFSRYPDH--M 79
Db 109 GHEFIFNAKLADVSAAEAASHL-----TTPVFGKGIYDPCNRLM 149
QY 80 KQHDFFKSAM-PEGYV-----QERTIFPKDGNKYKTRAEVKEGDTLVNRIELKGIDP 131
Db 150 EQKKFKVKGALTKFAFVRYPLIAEEIYKYFRNSKFKINENNSGIVDMVMVSOPEM--TTF 207
QY 132 KEDGNILGHKLEYNHSHNVYIMADQKNGIKVNFKIRHNIEDGVSQVLADHYQNTPIGD 191
Db 208 TARSLSLKGEMRKDLDTDFAYLSYDLDKGFTPINF--VFENPLPLEHYRKRDHAQQALS--- 263
QY 192 GPVLLPDNHYLSQSALSQKDPNEKRD 217
Db 264 -----GTYSMLIKERREKND 278

RESULT 10
D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN H10917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Spriggs D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.D., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RT Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U32773; AAC22575.1; --
DR TIGR; H10917; --
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac surface Ag; 1.
KW Antigen; Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 795 AA; 87478 MW; B85691FC22E6D44 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 795;
Best Local Similarity 21.9%; Pred. No. 15;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;
QY 65 LSYGVQCFSRYPDHMKQHDF-----FKSAMPEGYVOE-----RTI 99
Db 427 IGYGTESGISYQASVKQDNFLGTGAAGVSTAGTKNDYGTSTVNLGYTEPYFTTKDGVSLGGNV 486
QY 100 FFKDDGNKYKTRAEVKEGDTLVNRIELKGIDFKEDGNI--LCH-----KLEYNYS 148
Db 487 FFENVDSKSDTSSNYKRTTYGSNVTL-GFPVNNNSYVVGCHTYNKSINFALEYN--- 542
QY 149 HNVYIMADQK-NGIKVNFKIRHNIEDGVSQVLADHYQQ-----NTPIGDGFVLL 196
Db 543 RNLYIQSMKFKNGIKTN-----DPDFSGWYNLSNRGVYFTPKGVKASLG-GRVTI 593
QY 197 P--DNHYLSQALSQKDPNEKRDHVVLLGFTVTAAGITLG 233
Db 594 PGSDNKYKLSADVOGFYPLDRDLWVWSAKASAGYANG 632

RESULT 11
D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus influenzae.";
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Minna / Serotype B, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae species and may represent a universal protective antigen against invasive disease.";
RT Infect. Immun. 65:3701-3707(1997).
RL
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U13961; AAA85645.1; --
DR EMBL; U60832; AAB61974.1; --
DR EMBL; U60833; AAB61976.1; --
DR PIR; JC4078; JC4078.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac surface Ag; 1.
KW Antigen; Outer membrane; Signal; POTENTIAL.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE53896AF1B CRC64;
```

```
DR InterPro; IPR009008; Valrs_1Iers_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PRO0985; TRNASYNTHLEU.
DR TIGSFams; TIGR00396; leus_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "RMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99823 MW; 4C2EE01B8FDC497E CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 17;
Matches 44; Conservative 28; Mismatches 69; Indels 57; Gaps 10;

QY 50 TTGKLPVPWPTLVTLISYGVQCFSRYPDHMKQHDFFKSAFPGYVQERTIFFKDDGNY-- 107
DB 321 TNEQLPV-WVANFVLMAYGTGAVMAVPGHQDQDEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 108 -----KTRAEVKPEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVI 153
DB 378 STWEPDVWRDWTADKTR---EFE---LINSAEFDGLDYQGAPEVLAERFE----- 421
QY 154 MADKQKNG-IKYNFKIRHNIEDSGVOLADHYQOQNTPI-----GDGPFVLLPDN 199
DB 422 ---RQGRGQRRVNYRLR----DWGVSQRQYWGCPFVIYCTGCGAVPVPENQLPVILPEN 474
QY 200 -HYLSTQSALSADPNEKR 216
DB 475 VAFSGTGSPIKTDPEWRK 492

RESULT 13
VIT4_CAEL
ID VIT4_CAEL STANDARD; PRT; 1603 AA.
AC P18947; Q9BPP3;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
GN VIT-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T., Spieth J., Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=85269643; PubMed=4022780;
RA Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;
RT "The C. elegans vitellogenin genes: short sequence repeats in the
RT promoter regions and homology to the vertebrate genes.";
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -!- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
CC building the intestine of adult hermaphroditic individuals; they
CC are contranationally secreted into the body cavity and
CC subsequently taken up by the gonad.
CC -!- SIMILARITY: Contains 1 WVF domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```


CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR ENBL; M22056; AAA45738.1; -;
DR PIR; A30082; JDLVHH.
DR InterPro; IPR001462; DNAPol_viral_C.
DR InterPro; IPR00201; DNAPol_viral_N.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00336; DNA_pol_viral_C; 1.
DR Pfam; PF00242; DNA_pol_viral_N; 1.
DR Pfam; PF00078; rvt; 1.
DR ProDom; PD000814; DNAPol_viral_C; 1.
KW Transferrase; RNA-directed DNA polymerase; DNA-directed DNA polymerase;
KW Hydrolase; Nuclease; Endonuclease; DNA replication; DNA-binding.
SQ SEQUENCE 788 AA; 90070 MW; FB44F38F75EADF44 CRC64;

Query Match 6.7%; Score 85.5; DB 1; Length 788;
Best Local Similarity 19.7%; Pred. No. 18;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;
QY 58 WPTLVTLVSYGVQCFSRYPDHMKQH-----DPEKSAMPEGYVQERT-----IPFKDDGNYK 108
DB 139 WPKSISYLPVHSGVGPYKPEFQNHESLVNDYLNKLFEGILYKRVSKHLVTFK--GPYF 196
QY 109 T-----RAEVKFGDTLVNRIELKGIDFKEDGNILGHKLEYNVNSHNYIM 154
DB 197 TWEQKHLVPOQHGYASKINDROESRRRIITATSSRKNDSSRI-----FGAHN----- 245
QY 155 ADKQKNGIKVNFKIRHNIEDSGVOLADHYQ-----QNTPIGDGPFLL-PDNHYL 202
DB 246 -----NGRKISY---HSTRDGSRLSGRTSDPTSRGALAGDSTPIPGSTAHPSTHHV 297
QY 203 STQ-----SALSKDPNEKR 216
DB 298 DRRRQKGGVLOAISRPESETR 320

Search completed: June 21, 2004, 15:55:22
Job time : 7.55556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds
(without alignments)
2458.984 Million cell updates/sec

Title: US-09-887-784-222L
Perfect score: 1274
Sequence: 1 MVSKEELFTGVVILVELD.....VLGFVTAAGITLGMDELYK 239

Scoring table: BLOSOM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues 1017041

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTEMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1243	97.6	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1240	97.3	238	5 Q93125	Q93125 aequorea vi
3	1238	97.2	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1237	97.1	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1205	94.6	238	5 Q17105	Q17105 aequorea vi
6	1190	93.4	238	5 Q17106	Q17106 aequorea vi
7	1085	85.2	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1081	84.9	238	5 Q8WPC9	Q8wpc9 aequorea ma
9	1077	84.5	238	5 Q8WTC4	Q8wtc4 aequorea ma
10	1075	84.4	238	5 Q8WTD0	Q8wtd0 aequorea ma
11	1074	84.3	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1074	84.3	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1072	84.1	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1070	84.0	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5 Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5 Q7Z0W5	Q7z0w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	244.5	19.2	236	5	Q8T6U0	Q8t6u0 dendronephth
19	242.5	19.0	225	5	Q7Z0W9	Q7z0w9 montastraea
20	240	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyll
21	238.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q7Z0W4	Q7z0w4 montastraea
23	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
24	232	18.2	225	5	Q8T5F1	Q8t5f1 montastraea
25	214	16.8	227	5	Q7Z0W6	Q7z0w6 montastraea
26	214	16.8	234	5	Q7Z0W7	Q7z0w7 montastraea
27	212.5	16.7	259	5	Q8MMA2	Q8mma2 agaricia fr
28	212	16.6	239	5	Q8MMA1	Q8mma1 agaricia ag
29	209	16.4	234	5	Q8T5F2	Q8t5f2 montastraea
30	209	16.4	234	5	Q8MU47	Q8mu47 montastraea
31	208.5	16.4	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	206	16.2	227	5	Q962P9	Q962p9 montastraea
33	206	16.2	227	5	Q7Z0W8	Q7z0w8 montastraea
34	205.5	16.1	232	5	Q9GP15	Q9gp15 anemonia su
35	204.5	16.1	214	5	Q86LV7	Q86lv7 meandrina m
36	204	16.0	221	5	Q95P04	Q95p04 gonopora t
37	203.5	16.0	214	5	Q86LV8	Q86lv8 meandrina m
38	203.5	16.0	238	5	Q9BLV9	Q9bly9 renilla mue
39	202	15.9	227	5	Q95VT0	Q95vt0 montastraea
40	201.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
41	201.5	15.8	232	5	Q9GZ28	Q9gz28 anemonia su
42	198.5	15.6	222	5	Q7Z168	Q7z168 cerianthus
43	198.5	15.6	225	5	Q8T6T9	Q8t6t9 radianthus
44	198	15.5	235	5	Q8T5F0	Q8t5f0 scolymia cu
45	197.5	15.5	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	PRELIMINARY;	PRT;	238 AA.
ID	Q8GHE2		
AC	Q8GHE2;		
DT	01-MAR-2003 (Tremblrel. 23, Created)		
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)		
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)		
DE	Green fluorescence protein.		
DE	2289GFP.		
GN	Azotobacter vinelandii.		
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
OC	Pseudomonadaceae; Azotobacter.		
OX	NCBI_TaxID=354;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=DSM2289;		
RA	Koranyi P., Berenyi M., Burg K.;		
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria		
RL	Azomonas and Azotobacter."		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF324408; AAN86140.1; -		
DR	GO; GO:0006091; P:energy pathways; IEA.		
DR	InterPro; IPR0009017; GFP_like.		
DR	InterPro; IPR000786; Green_fl_protein.		
DR	Pfam; PF01353; GFP; 1.		
DR	PRINTS; PR01229; GFLUORSCENT.		
DR	ProDom; PD013756; Green_fl_protein; 1.		
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;		

Query Match 97.6%; Score 1243; DB 2; Length 238;
Best Local Similarity 97.9%; Pred. No. 1.9e-96;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy	2	VSKGELFTGVVPIVLVELDGVNGHKFSVSGEGEDATYKGLTKLKEICTTGKLPVWPPTL	61
Db	1	MKKGELFTGVVPIVLVELDGVNGHKFSVSGEGEDATYKGLTKLKEICTTGKLPVWPPTL	60
Qy	62	VTTLSYGVQCFRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNVYKTRAEVKEGDTLLV	121

```
Db 61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180
QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 238

RESULT 2
Q93125 Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-Optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Cormack B.P., Bertram G., Egerton M., Cow N.A.R., Falkow S.,
RA Brown A.U.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSP; P42212; IBBP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 28640 MW; A28622809A9DEA60 CRC64;

Query Match 97.3%; Score 1240; DB 5; Length 238;
Best Local Similarity 97.3%; Pred. No. 3.4e-96;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 60

Db 62 VTTLSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 238

RESULT 3
Q8GHE4 Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Green fluorescent protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 97.1%; Score 1237; DB 2; Length 238;
Best Local Similarity 97.5%; Pred. No. 5e-96;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 60

Db 62 VTTLSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 238

RESULT 4
Q8GHE3 Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Green fluorescent protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 97.1%; Score 1237; DB 2; Length 238;
Best Local Similarity 97.5%; Pred. No. 5e-96;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 60

Db 62 VTTLSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 238
```

Best Local Similarity 97.5%; Pred. No. 6.le-96;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKRHDFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLENYNSHNYIMADKQKGIKVNFKIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLENYNSHNYIMADKQKGIKVNFKIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
GN Green fluorescent protein (Fragment).
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -;
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.6%; Score 1205; DB 5; Length 238;
Best Local Similarity 94.1%; Pred. No. 3e-93;
Matches 224; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLENYNSHNYIMADKQKGIKVNFKIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLENYNSHNYIMADKQKGIKVNFKIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.

AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -;
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;

Query Match 93.4%; Score 1190; DB 5; Length 238;
Best Local Similarity 93.3%; Pred. No. 5.4e-92;
Matches 222; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLENYNSHNYIMADKQKGIKVNFKIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLENYNSHNYIMADKQKGIKVNFKIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -;
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP like.
DR Pfam; PF01353; GFP; 1.

```
DR PRINTS; PRO1229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 658FD75E88926903 CRC64;

Query Match      85.2%; Score 1085; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 3.6e-83;
Matches 196; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDGYKLTIKFICTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKAMPGEYQERTIFFKDDGNKYKTRAEVFEGLTLV 121
   .|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTLSYGIQCFARYPEHMKWDFKAMPGEYQERTIFFQDDGKYKTRGEVFEGLTLV 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGMDPKEDGNILGHKLEYNHNHYIMPDKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLIPINHLYSTQTAISKDRNETRDHMLVLEFFSACGHTGMDLYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qian Y.X., Pang S.Q.,
  Li S.J., Xia N.S.;
  Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013824; AAK02062.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;

Query Match      84.9%; Score 1081; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 7.7e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDGYKLTIKFICTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKAMPGEYQERTIFFKDDGNKYKTRAEVFEGLTLV 121
   .|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTFSYGIQCFARYPEHMKWDFKAMPGEYQERTIFFQDDGKYKTRGEVFEGLTLV 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGMDPKEDGNILGHKLEYNHNHYIMPDKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLIPINHLYSTQTAISKDRNETRDHMLVLEFFSACGHTGMDLYK 238
```

```
RESULT 9
Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
  Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.5%; Score 1077; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 1.7e-82;
Matches 197; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDGYKLTIKFICTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGVVPIVELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKAMPGEYQERTIFFKDDGNKYKTRAEVFEGLTLV 121
   .|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTLSYGIQCFARYPEHMKWDFKAMPGEYQERTIFFQDDGKYKTRGEVFEGLTLV 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGMDPKEDGNILGHKLEYNHNHYIMPDKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLIPINHLYSTQTAISKDRNETRDHMLVLEFFSACGHTGMDLYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qian Y.X., Pang S.Q.,
  Li S.J., Xia N.S.;
  Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013824; AAK02062.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;

Query Match      84.9%; Score 1081; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 7.7e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDGYKLTIKFICTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKAMPGEYQERTIFFKDDGNKYKTRAEVFEGLTLV 121
   .|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTFSYGIQCFARYPEHMKWDFKAMPGEYQERTIFFQDDGKYKTRGEVFEGLTLV 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGMDPKEDGNILGHKLEYNHNHYIMPDKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLIPINHLYSTQTAISKDRNETRDHMLVLEFFSACGHTGMDLYK 238
```



```

Db      1  MSKGEEELFGVWPILVELDGDVHGKXFSVRGEGEGDADYCKLEIKFICTTGGKLPVWPPTL  60
Qy      62  VTTLSVGVOCFSRYPDHMKQHDFFKSAPEGVYQERTIEFKDGNKYKTRAEVKFEGDTLV  121
Db      61  VTTLGVGILCFARYPEHMKNDFFKSAPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV  120
Qy      122  NRIELGIDIFKEDGNLTGHKLEYNVNSHNVIYIMADKKNGIKVNFKIRHNIEDGSVOLAD  181
Db      121  NRIELGMDIFKEDGNLTGHKLEYNFNSHNVIYIMPKDKNGLKVNFKIRHNIEGGVQLAD  180
Qy      182  HYQONTPIGDGPVLLPDNHYLSTQSALSOPNEKRDHVMULLGPFVTAAGITLGMDELYK  239
Db      181  HYQTNVPLGDGPVLLIPINHYLSYQTATSKDRNETRDHMFLEEFSCGHTHGMDELYK  238

RESULT 15
Q95UA7 PRELIMINARY; PRT; 225 AA.
ID AC Q95UA7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cyan fluorescent protein (Fragment).
OC Montastraea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Falkowski P.G., Sun Y.;
RT "Montastraea cavernosa fluorescent protein.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056460; AAL17905.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_LUOESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON TER 225
SQ SEQUENCE 225 AA; 25775 MW; 52DB2F716D083524 CRC64;

Query Match 19.8%; Score 252.5; DB 5; Length 225;
Best Local Similarity 31.1%; Pred. No. 2.8e-13;
Matches 64; Conservative 43; Mismatches 82; Indels 17; Gaps 7;

Qy      12  VVPIVELDGDVNGHKFSVSGEGEGDATYCKLTLLKF-ICTTGLKLPVWPPTLVLTLSYGVQ  70-
Db      7  VMKIKLAMDGIIVNGHKPMITGEGEGKPPGEGTHIILKVESGGLPFAYDILTTAFQYGNR  66
Qy      71  CFGRYPDHMKQHDFFKSAPEGVYQERTIEFKDGNKYKTRAEVKFEGDTLVNRIELGKID  130
Db      67  VFTKYPKDIQ-DFYKQSPFEGSWERSMTFEDQGCVCTVTSIDIKLEGDCFFYEIRFYGVN  124
Qy      131  FKEDGNILGHK-LEYNVNSHNVIYIMADKKNGIKVNFKIRHNIEDGSVOLADHYQONTPI  189
Db      125  FPSGSGPVNQKTLKWEPESTENMVV-----RDGVLLGDVSRITLLEGD----KHRCNFRS  175
Qy      190  GDGP---VLLPDNHYLSTQ-SALSCK  211
Db      176  TYGAKKGVLPEYHFDVHRIEILSHD  201

Search completed: June 21, 2004, 16:00:10
Job time : 30.7778 secs

```


XX The invention relates to a fluorescent protein derived from green
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations
 CC at F64L and E222G has a bigger compared to other GFP's making it very
 CC suitable for high throughput screening due to better resolution. The
 CC fluorescent protein is useful in *in vitro* assays for measuring protein
 CC kinase activity or dephosphorylation activity, or for measuring protein
 CC redistribution. The fluorescent protein is useful in studying cellular
 CC functions in living cells; as protein tags in transgenic animals, living
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.
 CC The fluorescent protein is also useful as a cell or organelle integrity
 CC marker, a marker for changes in cell morphology, as transfection marker,
 CC and as a marker to be used in combination with fluorescence activated
 CC cell sorting (FACS). The novel proteins can also be used as reporters to
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent
 CC protein is also useful as markers in transcriptional and translational
 CC fusions for performing transposon vector mutagenesis and as a reporter
 CC for bacterial detection. Transposons encoding the fluorescent protein are
 CC useful for screening promoters and for tagging plasmids and chromosomes.
 CC The fluorescent protein engineered into the genome of a phage is useful
 CC for designing diagnostic tool. The present sequence is a DNA encoding
 CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant
 XX
 SQ Sequence 239 AA;

Query Match 99.8%; Score 1271; DB 5; Length 239;
 Best Local Similarity 99.8%; Pred. No. 1.1e-122;
 Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 2
 ABR40352
 ID ABR40352 standard; protein; 363 AA.

XX ABR40352;
 XX
 XX 08-JUL-2003 (first entry)
 DT Human amino acid sequence SEQ ID NO: 6.
 DE Human; heterologous conjugate; intracellular protein.

XX Homo sapiens.
 OS Aequoria victoria.
 XX WO2003029827-A2.
 XX 10-APR-2003.

XX 01-OCT-2002; 2002WO-DK000651.
 XX 01-OCT-2001; 2001DK-00001433.
 XX 11-OCT-2001; 2001US-032896P.
 XX (BIOL-) BIOLMAGE AS.
 XX Terry BR, Nielsen SJ;

XX WPI: 2003-430211/40.
 DR N-PSDB; ACC72604.
 XX Novel cell for identifying modulators of protein interaction, contains a
 PT first conjugate comprising anchor protein, second conjugate having type B
 PT interactor protein and third conjugate with detectable group.
 XX Disclosure; Page 112-113; 118pp; English.

XX The invention relates to a novel cell, comprising three heterologous
 CC conjugates (HC), a first HC (HC1) comprising an anchor protein that
 CC specifically binds to an internal structure within the cell conjugated to
 CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
 CC type B conjugated to a first protein of interest, and a third HC (HC3)
 CC comprising a second protein of interest conjugated to detectable group.
 CC The cell is useful for detecting if a compound disrupts or induces the
 CC interaction between two intracellular proteins. The cell is also useful
 CC for screening compounds that modulate the interaction between two
 CC intracellular proteins. The present sequence is used in the
 CC exemplification of the invention
 XX
 SQ Sequence 363 AA;

Query Match 99.8%; Score 1271; DB 6; Length 363;
 Best Local Similarity 99.6%; Pred. No. 2e-122;
 Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60
 DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRLEKLGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
 DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 3
 AAG65781
 ID AAG65781 standard; protein; 893 AA.

XX AAG65781;
 XX
 XX 07-JAN-2002 (first entry)
 DT Amino acid sequence of HSPDB4A1-E222G fusion protein.
 DE PDE4; central nervous system; antinflammatory; cytostatic; nootropic;
 KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
 KW fusion protein.

XX Homo sapiens.
 OS Aequorea victoria.
 XX WO200179526-A2.
 XX 25-OCT-2001.

XX 11-APR-2001; 2001WO-DK000264.
 XX 17-APR-2000; 2000DK-00000651.
 XX 29-MAY-2000; 2000DK-00000849.
 XX (BIOI-) BIOIMAGE AS.

XX Terry BR, Nielsen SJ;

PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
DR WPI; 2001-611727/70.
DR N-PSDB; AAI66852.
XX
PT Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
PS Example 1; Page 156-160; 160pp; English.
XX
CC The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
CC protein
XX
SQ Sequence 893 AA;

Query Match 99.8%; Score 1271; DB 4; Length 893;
Best Local Similarity 99.6%; Pred. No. 7.7e-122;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTIKFICTTGKLPVPWPT 60
DB 655 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTIKFICTTGKLPVPWPT 714

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 774

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 834

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 893

RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA.
XX
AC AAG65782;
XX
DT 07-JAN-2002 (first entry)
XX
DE Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
KW PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW fusion protein.
XX
OS Homo sapiens.
XX Aequorea victoria.
XX
PN WO200179526-A2.
XX

PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-DK000264.
XX
PR 17-APR-2000; 2000DK-00000651.
PR 29-MAY-2000; 2000DK-00000849.
XX
PA (BIOI-) BIOIMAGE AS.
XX
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66853.
XX
PT Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
PS Example 1; Page 162-167; 160pp; English.
XX
CC The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
CC protein
XX
SQ Sequence 1132 AA;

Query Match 99.8%; Score 1271; DB 4; Length 1132;
Best Local Similarity 99.6%; Pred. No. 1.1e-121;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTIKFICTTGKLPVPWPT 60
DB 894 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTIKFICTTGKLPVPWPT 953

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 1132

RESULT 5
AAE17517
ID AAE17517 standard; protein; 239 AA.
XX
AC AAE17517;
XX
DT 22-APR-2002 (first entry)
XX
DE Enhanced F64L jellyfish green fluorescent protein mutant.
XX

KW	Jellyfish; green fluorescent protein; GFP; protein redistribution;	
KW	cellular function; genetic reporter; mutant; Stoke's shift; mutain.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 65	/notes= "Wild type Phe substituted with Leu; This
FT		corresponds to position 64 in the wild type protein"
XX		
XX		
PN	WO200198338-A2.	
XX		
PD	27-DEC-2001.	
XX		
XX	18-JUN-2001; 2001WO-EP006848.	
XX		
PR	19-JUN-2000; 2000DK-00000953.	
PR	20-JUN-2000; 2000US-0212681P.	
PR	10-MAY-2001; 2001DK-00000739.	
PR	10-MAY-2001; 2001US-0290170P.	
XX		
PA	(BIOI-) BIOIMAGE AS.	
XX		
PI	Bjorn SP, Pagliaro L, Thastrup O;	
XX		
DR	WPI; 2002-098224/13.	
DR	N-PSDB; AAD28162.	
XX		
PT	Novel fluorescent protein in in vitro assay for measuring protein kinase	
PT	activity or dephosphorylation activity, or for measuring protein	
PT	redistribution, has a green fluorescent protein with F64L and E222G	
PT	mutation.	
XX		
PS	Example 1; Page 35; 41pp; English.	
XX		
CC	The invention relates to a fluorescent protein derived from green	
CC	fluorescent protein (GFP) or its analogue. The GFP containing mutations	
CC	at F64L and E222G has a bigger compared to other GFP's making it very	
CC	suitable for high throughput screening due to better resolution. The	
CC	fluorescent protein is useful in invitro assays for measuring protein	
CC	kinase activity or dephosphorylation activity, or for measuring protein	
CC	redistribution. The fluorescent protein is useful in studying cellular	
CC	functions in living cells; as protein tags in transgenic animals, living	
CC	and fixed cells; organelle tags, secretion marker and genetic reporter.	
CC	The fluorescent protein is also useful as a cell or organelle integrity	
CC	marker, a marker for changes in cell morphology, as transfection marker,	
CC	and as a marker to be used in combination with fluorescence activated	
CC	cell sorting (FACS). The novel proteins can also be used as reporters to	
CC	monitor live or dead biomass of organisms, such as fungi. The fluorescent	
CC	protein is also useful as markers in transcriptional and translational	
CC	fusions for performing transposon vector mutagenesis and as a reporter	
CC	for bacterial detection. Transposons encoding the fluorescent protein are	
CC	useful for screening promoters and for tagging plasmids and chromosomes.	
CC	The fluorescent protein engineered into the genome of a phage is useful	
CC	for designing diagnostic tool. The present sequence is enhanced F64L	
CC	Jellyfish green fluorescent protein (GFP) mutant	
XX		
SQ	Sequence 239 AA;	
	Query Match 99.1%; Score 1263; DB 5; Length 239;	
	Best Local Similarity 99.2%; Pred. No. 7.4e-122;	
	Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	
Qy	1 MVSKEELFTGVPILVELDGVNKHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60	
Db	1 MVSKEELFTGVPILVELDGVNKHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60	
Qy	61 LVTTLSGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNVYKTRAEVKEGDITL 120	
Db	61 LVTTLSGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNVYKTRAEVKEGDITL 120	
Qy	121 VNRTELKGDIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180	

Db	121 VNRTELKGDIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180	
Qy	181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239	
Db	181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239	
RESULT 6		
AAB22882		
ID	AAB22882 standard; protein; 239 AA.	
XX		
AC	AAB22882;	
XX		
DT	10-JAN-2001 (first entry)	
XX		
DE	Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.	
XX		
KW	Bioreactor protein; fusion protein; recognition site;	
KW	cellular targeting sequence; cellular localisation; fluorescent protein;	
KW	protease activity detection; toxin detection; cellular stress detection;	
KW	drug discovery; cell based screening.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
PN	WO200050872-A2.	
XX		
PD	31-AUG-2000.	
XX		
XX	25-FEB-2000; 2000WO-US004794.	
PR	26-FEB-1999; 99US-0122152P.	
PR	08-MAR-1999; 99US-0123399P.	
PR	12-JUL-1999; 99US-00352171.	
XX	(CELL-) CELLOMICS INC.	
PA	Giuliano KA, Kapur R;	
PI		
XX	WPI; 2000-594086/56.	
DR	N-PSDB; AAA93373.	
XX		
PT	Automated cell-based characterization of toxin by contacting cells	
PT	containing luminescent reporter molecules with test substance and	
PT	analyzing optically.	
XX		
PS	Example 11; Fig 29A; 336pp; English.	
XX		
CC	The invention relates to systems, methods and reagents for cell-based	
CC	screening or detection of compounds which affect particular biological	
CC	functions. The methods of the invention utilise fluorescent bioreactor	
CC	molecules which, when acted on by a compound of interest, cause an	
CC	alteration in the cellular distribution of at least the fluorescent	
CC	moiety. In one embodiment, the biosensors comprise heat shock proteins	
CC	(HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent	
CC	protein (GFP), or derivatives thereof). Such biosensors are located in	
CC	the cytoplasm, but on stress activation translocate to the nucleus. In	
CC	another embodiment bioreactor proteins can be used to detect protease	
CC	activity. Such protease bioreactor proteins comprise one or more	
CC	fluorescent proteins; a recognition signal which is cleaved by the	
CC	protease; and at least one cellular localisation signal. The latter two	
CC	components may be components of a single protein which is acted upon by	
CC	the protease, or may be from heterologous sources. Due to the	
CC	localisation signal, the bioreactor protein is localised to a particular	
CC	region of the cell, once acted on by the protease of interest, the	
CC	fluorescent protein is cleaved from the localisation sequence, and is	
CC	free to migrate to other locations within the cell. The presence of a	
CC	second localisation signal attached to the fluorescent protein enables	
CC	the fluorescent protein to be directed to a different cellular	
CC	compartment after cleavage of the protease recognition sequence. The	
CC	change in distribution of the fluorescent protein can be detected using	
CC	imaging methods with a high degree of spatial resolution. The methods and	

CC biosensors of the invention can be used to investigate a wide range of
CC cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased quantity
CC and speed of data collection and faster evaluation of drug candidates.
CC Sequences AAB2881-82885 represent fluorescent proteins which may be used
CC as components of biosensor fusion proteins of the invention
XX
XX Sequence 239 AA;

Query Match 98.9%; Score 1260; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-121;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
QY 61 LVTTLGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 7
AAV54349
ID AAY54349 standard; protein; 239 AA.
XX
AC AAY54349;
XX
DT 06-APR-2000 (first entry)
XX
DE Amino acid sequence of the mutant green fluorescent protein EGFP.
XX
KW Fluorescent protein; green fluorescent protein; emission intensity;
KW fluorescence; pH detection; pH sensor; EGFP.
XX
OS Synthetic.
OS Aequorea victoria.
XX
FH Key Location/Qualifiers
FT Misc-difference 65 /note= "wild type Phe substituted with Leu"
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"
FT Misc-difference 232 /note= "wild type His substituted with Leu"
FT
XX
PN WO964592-A2.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US012850.
XX
PR 09-JUN-1998; 98US-00094359.
PR 13-OCT-1998; 98US-00172063.
XX
PA (REGC) UNIV CALIFORNIA.
PA (UYOR-) UNIV OREGON STATE.
XX

PI Tsien RY, Llopis J, Wachter RM;
XX WPI; 2000-116540/10.
DR N-PSDB; AAZ45642.
XX
PT New functional engineered green fluorescent proteins, used for measuring
the pH in biological samples and cells.
XX
PS Disclosure; Page 9; 89pp; English.
XX
CC The present sequence represents a functional engineered fluorescent
protein based on the Aequorea green fluorescent protein (GFP). The
CC emission intensity changes as pH varies between 5 and 10 of the present
protein are novel. The functional engineered fluorescent proteins show
CC reversible changes in fluorescence over physiological pH ranges. They can
be used for determining the pH of samples and cells. The polynucleotides
CC can also be used to produce transgenic animals. The fluorescent protein
pH sensors can be delivered to cells in the form of polynucleotides
CC encoding the protein sensor fused to a targeting signal. The targeting
signal directs the expression of the protein sensors to restricted cell
CC locations. This makes it possible to measure the pH of a precisely
defined cellular region or organelle
XX
XX Sequence 239 AA;

Query Match 98.9%; Score 1260; DB 3; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-121;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60
QY 61 LVTTLGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 8
AAV79584
ID AAY79584 standard; peptide; 239 AA.
XX
AC AAY79584;
XX
DT 29-AUG-2000 (first entry)
XX
DE EGFP signal domain.
XX
KW Protease; biosensor; EGFP; signal peptide; cell screening; assay;
KW analysis; drug discovery.
XX
OS Unidentified.
XX
PN WO200026408-A2.
XX
PD 11-MAY-2000..
XX
PF 29-OCT-1999; 99WO-US025431.
XX
PR 30-OCT-1998; 98US-0106308P.
PR 26-MAY-1999; 99US-0136078P.
XX
PA (CELL-) CELLOMICS INC.
XX
PI Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX WPI: 2000-365644/31.
 DR N-PSDB; AA27573.
 XX
 PT Recombinant nucleic acid encoding a protease biosensor useful for
 PT fluorescence based cell and molecular biochemical assays for drug
 PT discovery comprising three operably linked nucleic acid sequences.
 XX
 PS Claim 14; Fig 29A; 218pp; English.
 XX
 CC The present sequence is that of the EGFP signal domain, which can be
 CC included in novel recombinant protease biosensors (PBs) of the invention.
 CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)
 CC comprising at least 1 detectable polypeptide signal such as the present
 CC sequence; a second domain (see AAY79588-622) comprising at least 1
 CC protease recognition site; and a third domain (see AAY79623-37)
 CC comprising at least 1 reactant target sequence. A recombinant nucleic
 CC acid (see AAA27627-43) encoding the PB, an expression vector, and a
 CC genetically engineered host cell are also claimed. A claimed method for
 CC identifying compounds that modify protease activity in a cell involves
 CC contacting a host cell that possesses the recombinant PB with a test
 CC compound, and determining the PB distribution in the host cell, where
 CC changes in the distribution of the PB are correlated with modification of
 CC protease activity by the test compound. Claimed kits for identifying
 CC compounds that modify protease activity in a host cell include the
 CC recombinant nucleic acid, or the recombinant PB, or the vector, or the
 CC host cell. The PB is useful in high content screens to detect in vivo
 CC activation of enzymatic activity, and to identify specific activity based
 CC on cleavage of a known recognition motif
 XX
 SQ Sequence 239 AA;

Query Match 98.9%; Score 1260; DB 3; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.5e-121;
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVWPWT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVWPWT 60
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTGYVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 9
 AAB50804
 ID AAB50804 standard; protein; 239 AA.
 XX
 AC AAB50804;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Jellyfish GFP mutant EGFP.
 XX
 KW Aequorea victoria; jellyfish; fluorescent protein indicator;
 KW green fluorescent protein; GFP; linker moiety; sensor;
 KW calmodulin-binding domain; mutant; mutein.
 XX
 OS Aequorea victoria.
 XX
 PN WO200071565-A2.
 XX
 PD 30-NOV-2000.
 XX

PF 17-MAY-2000; 2000WO-US013684.
 XX
 PR 21-MAY-1999; 99US-00316919.
 PR 21-MAY-1999; 99US-00316920.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Tsien RY, Baird GA;
 XX
 DR WPI; 2001-032017/04.
 DR N-PSDB; AAC90488.
 XX
 PT Novel fluorescent proteins comprising a sensor protein inserted into
 PT them, useful for measuring the response of a sensor biological, chemical,
 PT electrical or physiological parameter in vivo or in vitro.
 XX
 XX Disclosure; Page 24; 94pp; English.
 XX
 CC The present sequence is a fluorescent protein used in the construction of
 CC a fluorescent protein indicator. The indicator comprises a sensor
 CC polypeptide that is responsive to a chemical, biological, electrical or
 CC physiological parameter, and a fluorescence protein functional group. The
 CC sensor polypeptide is operatively inserted into the fluorescent moiety.
 CC The fluorescent indicator is useful for detecting the presence of a
 CC response inducing member in a sample. The method involves contacting the
 CC sample with the indicator and detecting a change in fluorescence, in
 CC which a change is indicative of the effect of the parameter on the sensor
 CC polypeptide. The novel fluorescent proteins are advantageous due to their
 CC reduced size as compared to the FRET (fluorescence resonance energy
 CC transfer)-based sensors
 XX
 SQ Sequence 239 AA;

Query Match 98.9%; Score 1260; DB 4; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.5e-121;
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVWPWT 60
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVWPWT 60
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 DB 61 LVTTLTGYVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 10
 AAB85900
 ID AAB85900 standard; protein; 239 AA.
 XX
 AC AAB85900;
 XX
 DT 30-NOV-2001 (first entry)
 XX
 DE A. victoria green fluorescent protein (GFP) and linker sequence.
 XX
 KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
 KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;
 KW green fluorescent protein.
 XX
 OS Synthetic.
 OS Aequorea victoria.
 XX
 PN WO200168706-A1.
 XX

```
PD 20-SEP-2001.
XX
XX 14-MAR-2001; 2001WO-US008071.
XX
XX 15-MAR-2000; 2000US-0189698P.
XX
XX (MERI ) MERCK & CO INC.
XX
XX Marsh DJ;
XX
XX WPI; 2001-565791/63.
DR N-PSDB; AAH47304.
XX
XX Fusion proteins comprising melanin concentrating hormone receptor
PT peptides and fluorescent proteins, useful for identifying appetite
PT stimulants.
XX
XX Claim 2; Page 14; 71pp; English.
XX
XX The invention provides melanin concentrating hormone (MCH) receptor
CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
CC MCHR polypeptide regions from different species. The MCHR fusion protein
CC comprise MCHR polypeptide region and a fluorescent polypeptide region
CC joined directly, or via a linker, to the carboxy side of the MCHR
CC polypeptide region. The MCHR fusion proteins can be expressed by standard
CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
CC regulation of MCH activity stimulates food intake. The present sequence
CC represents a A. victoria green fluorescent protein (GFP) and a linker
CC sequence
XX
XX Sequence 239 AA;

Query Match      98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-121;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 11
AAB31171
ID AAB31171 standard; protein; 239 AA.
XX
XX AAB31171;
AC
XX
XX 02-APR-2001 (first entry)
DT
XX
XX Amino acid sequence of a green fluorescent protein (GFP).
DE
XX
XX Growth rate; death rate; reporter gene; luminescent protein; GFP.
KW fluorescent product; luciferase; green fluorescent protein; GFP.
XX
XX Aequorea victoria.
OS
XX
XX WO200075367-A1.
PN
XX
XX 14-DEC-2000.
PD
XX
XX 07-JUN-2000; 2000WO-FI000507.
PF

XX 07-JUN-1999; 99FI-00001296.
XX
XX (LILI/) LILIUS E.
PA (VIRT/) VIRT M.
XX
XX Lilius E, Virta M;
PI
XX WPI; 2001-061737/07.
DR N-PSDB; AAC86954.
XX
XX Assessing growth and death rates of a micro-organism in a desired
PT environment, by introducing 2 reporter genes encoding luminescent and
PT fluorescent products and detecting luminescent fluorescence.
XX
XX Disclosure; Page 27; 32pp; English.
XX
XX The specification describes a method for assessing the growth rate and
CC death rate of a micro-organism within a predetermined time period in a
CC desired environment. The method comprises introducing at least two
CC reporter genes encoding luminescent and/or fluorescent products into the
CC micro-organisms, incubating the micro-organism within the desired
CC environment, and detecting luminescence and/or fluorescence after a
CC predetermined time period. Use of two different markers within a micro-
CC organism enables the differentiation between growth and death rates. The
CC method is used to assess the growth rate and death rate of a micro-
CC organism within a predetermined time period in a desired environment. The
CC present sequence represents a green fluorescent protein (GFP), and is
CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
CC and is used in the method of the invention
XX
XX Sequence 239 AA;

Query Match      98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-121;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 12
AAG66198
ID AAG66198 standard; protein; 239 AA.
XX
XX AAG66198;
AC
XX
XX 17-JUN-2002 (first entry)
DT
XX
XX A. victoria green fluorescent protein (EGFP).
DE
XX
XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
KW green fluorescent protein; EGFP.
XX
XX Aequorea victoria.
OS
XX
XX JF2002045189-A.
PN
XX
XX 12-FEB-2002.
PD
XX
```


XX FH Key Location/Qualifiers
 FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"
 FT FT
 FT FT
 FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"
 FT FT
 FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"
 FT FT
 XX EP1178109-A1.
 XX XX
 XX PD 06-FEB-2002.
 XX XX
 XX 03-AUG-2001; 2001EP-00306650.
 XX XX
 PR 04-AUG-2000; 2000JP-00237166.
 XX XX
 XX (RIKE) RIKEN KK.
 XX XX
 XX Miyawaki A, Sawano A;
 XX WPI; 2002-208112/27.
 DR N-PSDB; AAD27910.
 DR XX
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.
 PT PT
 PT PT
 XX XX
 PS Example 1; Page 13-14; 31pp; English.
 XX XX
 CC The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria
 XX SQ Sequence 239 AA;
 Query Match 98.9%; Score 1260; DB 5; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.5e-121;
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MVSGBELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 DB 1 MVSGBELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
 DB 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRHDHVMVLFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRHDHVMVLFVTAAGITLGMDELYK 239

QY 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRHDHVMVLFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRHDHVMVLFVTAAGITLGMDELYK 239
 RESULT 15
 AAE34958
 ID AAE34958 standard; protein; 239 AA.
 XX AAE34958;
 AC AAE34958;
 XX DT 28-MAY-2003 (first entry)
 XX DE Aequorea victoria enhanced green fluorescent protein (EGFP).
 XX KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;
 KW kinase; enhanced green fluorescent protein; EGFP.
 XX OS Aequorea victoria.
 XX PN WO200295058-A2.
 XX PD 28-NOV-2002.
 XX 24-MAY-2002; 2002WO-US016955.
 PF 24-MAY-2001; 2001US-00865291.
 PR (REGC) UNIV CALIFORNIA.
 XX Tsien RY, Ting AY, Zhang J;
 WPI; 2003-148474/14.
 DR N-PSDB; AAD53428.
 XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphaminoacid binding domain, and acceptor molecule, in operative linkage.
 PT Disclosure; Col 56-57; 38pp; English.
 XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or operative linkage, a donor molecule, a phosphorylatable domain, a phosphaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention
 XX SQ Sequence 239 AA;
 Query Match 98.9%; Score 1260; DB 6; Length 239;
 Best Local Similarity 98.7%; Pred. No. 1.5e-121;
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MVSGBELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 DB 1 MVSGBELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
 DB 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
 QY 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRHDHVMVLFVTAAGITLGMDELYK 239
 DB 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRHDHVMVLFVTAAGITLGMDELYK 239

Db 181 DHYQONTPIGDPVLLPDNNHYLSTQSALSXDPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:12
Job time : 48.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds
(without alignments)
965.630 Million cell updates/sec

Title: US-09-887-784-222V

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVPIVELD.....VLGVFTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
 - 2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
 - 3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
 - 4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
 - 5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*
 - 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1260	98.9	239	3	US-09-172-063-3
2	1260	98.9	239	4	US-09-513-783A-46
3	1260	98.9	239	4	US-09-316-919-4
4	1260	98.9	239	4	US-09-602-641-3
5	1260	98.9	239	4	US-09-920-922-2
6	1260	98.9	281	3	US-09-062-102-1
7	1260	98.9	281	4	US-09-364-946-1
8	1260	98.9	294	4	US-09-513-783A-2
9	1260	98.9	323	3	US-09-172-063-21
10	1260	98.9	323	4	US-09-602-641-21
11	1260	98.9	354	3	US-09-085-305-6
12	1260	98.9	379	4	US-09-417-197-129
13	1260	98.9	434	4	US-09-800-170-48
14	1260	98.9	442	4	US-09-417-197-127
15	1260	98.9	459	4	US-09-513-783A-170
16	1260	98.9	544	4	US-09-417-197-113
17	1260	98.9	544	4	US-09-417-197-115
18	1260	98.9	604	4	US-09-417-197-59
19	1260	98.9	605	4	US-09-417-197-41
20	1260	98.9	606	4	US-09-417-197-65
21	1260	98.9	607	4	US-09-417-197-47
22	1260	98.9	630	4	US-09-417-197-63
23	1260	98.9	631	4	US-09-417-197-39
24	1260	98.9	633	4	US-09-417-197-45
25	1260	98.9	635	4	US-09-417-197-125
26	1260	98.9	642	2	US-08-818-253-2
27	1260	98.9	642	2	US-08-818-253-6

28	1260	98.9	642	3	US-08-818-252-2	Sequence 2, Appli
29	1260	98.9	642	3	US-08-818-252-6	Sequence 6, Appli
30	1260	98.9	652	2	US-08-818-253-4	Sequence 4, Appli
31	1260	98.9	652	3	US-08-818-252-4	Sequence 4, Appli
32	1260	98.9	718	4	US-09-417-197-75	Sequence 75, Appli
33	1260	98.9	719	4	US-09-417-197-51	Sequence 51, Appli
34	1260	98.9	726	4	US-09-417-197-71	Sequence 71, Appli
35	1260	98.9	727	4	US-09-417-197-139	Sequence 139, App
36	1260	98.9	783	4	US-09-513-783A-176	Sequence 176, App
37	1260	98.9	797	4	US-09-417-197-141	Sequence 141, App
38	1260	98.9	797	4	US-09-417-197-143	Sequence 143, App
39	1260	98.9	798	4	US-09-417-197-77	Sequence 77, Appli
40	1260	98.9	805	4	US-09-513-783A-178	Sequence 178, App
41	1260	98.9	806	4	US-09-417-197-53	Sequence 53, Appli
42	1260	98.9	836	4	US-09-417-197-61	Sequence 61, Appli
43	1260	98.9	842	4	US-09-417-197-43	Sequence 43, Appli
44	1260	98.9	843	4	US-09-417-197-117	Sequence 117, App
45	1260	98.9	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1
US-09-172-063-3
; Sequence 3, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Liopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-172-063-3

Query Match	98.9%;	Score	1260;	DB 3;	Length	239;			
Best Local Similarity	98.7%;	Pred. No.	1.5e-127;						
Matches	236;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;
QY	1	MVSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDATYVKLTAKFICTTGTGKLPVPWPT	60						
Db	1	MVSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDATYVKLTAKFICTTGTGKLPVPWPT	60						
QY	.61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNVKTAEVKFEGDTL	120						
Db	61	LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNVKTAEVKFEGDTL	120						
QY	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA	180						
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA	180						
QY	181	DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMYLGVFTAAAGITLGMDELYK	239						
Db	181	DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMYLGVFTAAAGITLGMDELYK	239						

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          98.9%; Score 1260; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

```

```

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CIP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          98.9%; Score 1260; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match          98.9%; Score 1260; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 2.1e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60

```

Db 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLEICTTGKLPVWPWT 60
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNVKTAEVKFEGDTL 120
Db 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNVKTAEVKFEGDTL 120
QY 121 VNRIELKIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 121 VNRIELKIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 9

US-09-172-063-21
; Sequence 21, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 09/094,359
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-172-063-21

Query Match 98.9%; Score 1260; DB 3; Length 323;
Best Local Similarity 98.7%; Pred. No. 2.4e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLEICTTGKLPVWPWT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLEICTTGKLPVWPWT 144
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNVKTAEVKFEGDTL 120
Db 145 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNVKTAEVKFEGDTL 204
QY 121 VNRIELKIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 205 VNRIELKIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLFVTAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21
; Sequence 21, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-602-641-21

Query Match 98.9%; Score 1260; DB 4; Length 323;
Best Local Similarity 98.7%; Pred. No. 2.4e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLEICTTGKLPVWPWT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTLEICTTGKLPVWPWT 144
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNVKTAEVKFEGDTL 120
Db 145 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNVKTAEVKFEGDTL 204
QY 121 VNRIELKIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
Db 205 VNRIELKIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLFVTAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6
; Sequence 6, Application US/09085305
; Patent No. 6191269
; GENERAL INFORMATION:
; APPLICANT: Pollock, Allan
; APPLICANT: Lovett, David H.
; APPLICANT: Turck, Johanna
; TITLE OF INVENTION: Selective Induction of Apoptosis in
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,305
; FILING DATE: 29-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Francis, Carol L
; REGISTRATION NUMBER: 36,513
; REFERENCE/DOCKET NUMBER: 6510/102US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-085-305-6

Query Match      98.9%; Score 1260; DB 3; Length 364;
Best Local Similarity 98.7%; Pred. No. 2.8e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
Db      126 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 185

QY      61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db      186 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 245

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db      246 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 305

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db      306 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 364

RESULT 12
US-09-417-197-129
; Sequence 129, Application US/09417197
; Patent No. 6518021
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP, et al.
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An I
; FILE REFERENCE: 3759-0110P
; CURRENT APPLICATION NUMBER: US/09/417,197
; CURRENT FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 129
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: actin-binding-domain-EGFP fusion
US-09-417-197-129

Query Match      98.9%; Score 1260; DB 4; Length 379;
Best Local Similarity 98.7%; Pred. No. 3e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
Db      141 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 200

QY      61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db      201 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 260

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db      261 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 320

; ATTORNEY/AGENT INFORMATION:
; NAME: Francis, Carol L
; REGISTRATION NUMBER: 36,513
; REFERENCE/DOCKET NUMBER: 6510/102US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-085-305-6

Query Match      98.9%; Score 1260; DB 3; Length 364;
Best Local Similarity 98.7%; Pred. No. 2.8e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
Db      126 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 185

QY      61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db      186 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 245

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db      246 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 305

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db      306 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 364

RESULT 12
US-09-417-197-129
; Sequence 129, Application US/09417197
; Patent No. 6518021
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP, et al.
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An I
; FILE REFERENCE: 3759-0110P
; CURRENT APPLICATION NUMBER: US/09/417,197
; CURRENT FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 129
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: actin-binding-domain-EGFP fusion
US-09-417-197-129

Query Match      98.9%; Score 1260; DB 4; Length 379;
Best Local Similarity 98.7%; Pred. No. 3e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
Db      141 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 200

QY      61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db      201 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 260

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db      261 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 320

; APPLICANT: KinSELLa, Todd
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/800,170
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Synechocystis PCC6803
US-09-800-170-48

Query Match      98.9%; Score 1260; DB 4; Length 434;
Best Local Similarity 98.7%; Pred. No. 3.7e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
Db      196 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 255

QY      61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db      256 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQVQERTIFFKDDGNGYKTRAEVKFEGDTL 315

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180
Db      316 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 375

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db      376 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 434

RESULT 14
US-09-417-197-127
; Sequence 127, Application US/09417197
; Patent No. 6518021
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP, et al.
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An I
; FILE REFERENCE: 3759-0110P
; CURRENT APPLICATION NUMBER: US/09/417,197
; CURRENT FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 127
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-RhoA fusion
US-09-417-197-127

Query Match      98.9%; Score 1260; DB 4; Length 442;
Best Local Similarity 98.7%; Pred. No. 3.8e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 MYSKGEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
```

```
Db      1  MVSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPVWPPT 60
Qy      61  LVTTLTSGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db      61  LVTTLTSGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Qy     181  DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGVFTAAGITLGMDELYK 239
Db     181  DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGVFTAAGITLGMDELYK 239
```

```
RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170
```

```
Query Match      98.9%; Score 1260; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 4e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1  MVSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPVWPPT 60
Db      1  MVSKGELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPVWPPT 60
Qy     61  LVTTLTSGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db     61  LVTTLTSGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Qy     181  DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGVFTAAGITLGMDELYK 239
Db     181  DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGVFTAAGITLGMDELYK 239
```

Search completed: June 21, 2004, 16:04:04
Job time : 13.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds
(without alignments)
1940.117 Million cell updates/sec

Title: US-09-887-784-222V

Perfect score: 1274

Sequence: 1 MWSKGELFTGVVPIVELD.....VLGVFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1271	99.8	239	9	US-09-887-784-4
2	1271	99.8	239	12	US-10-296-953-4
3	1271	99.8	363	14	US-10-270-223-6
4	1271	99.8	893	14	US-10-257-909A-30
5	1271	99.8	1132	14	US-10-257-909A-32
6	1263	99.1	239	9	US-09-887-784-2
7	1263	99.1	239	12	US-10-296-953-2
8	1260	98.9	239	9	US-09-920-922-2
9	1260	98.9	239	9	US-09-999-745-4
10	1260	98.9	239	10	US-09-866-538-4
11	1260	98.9	239	10	US-09-797-496B-2
12	1260	98.9	239	10	US-09-794-308-4
13	1260	98.9	239	10	US-09-865-291-4
14	1260	98.9	239	12	US-10-457-982-3
15	1260	98.9	239	14	US-10-121-258-13

16	1260	98.9	239	14	US-10-221-461-7	Sequence 7, Appli
17	1260	98.9	239	14	US-10-100-957A-46	Sequence 46, Appl
18	1260	98.9	239	14	US-10-177-390-2	Sequence 2, Appli
19	1260	98.9	239	14	US-10-338-411-3	Sequence 3, Appli
20	1260	98.9	239	15	US-10-370-570-4	Sequence 4, Appli
21	1260	98.9	239	15	US-10-389-640-3	Sequence 3, Appli
22	1260	98.9	259	12	US-09-931-232-1	Sequence 1, Appli
23	1260	98.9	281	12	US-10-314-861-11	Sequence 37, Appl
24	1260	98.9	288	14	US-10-314-861-37	Sequence 35, Appl
25	1260	98.9	293	14	US-10-314-861-35	Sequence 2, Appli
26	1260	98.9	294	14	US-10-100-957A-2	Sequence 39, Appl
27	1260	98.9	295	14	US-10-314-861-39	Sequence 33, Appl
28	1260	98.9	299	14	US-10-314-861-33	Sequence 31, Appl
29	1260	98.9	305	14	US-10-314-861-31	Sequence 35, Appl
30	1260	98.9	308	14	US-10-033-717-35	Sequence 29, Appl
31	1260	98.9	311	14	US-10-314-861-29	Sequence 11, Appl
32	1260	98.9	320	14	US-10-338-411-11	Sequence 7, Appli
33	1260	98.9	320	15	US-10-389-640-11	Sequence 13, Appl
34	1260	98.9	323	12	US-10-457-982-21	Sequence 21, Appl
35	1260	98.9	323	14	US-10-338-411-7	Sequence 7, Appli
36	1260	98.9	323	14	US-10-338-411-13	Sequence 13, Appl
37	1260	98.9	323	15	US-10-389-640-7	Sequence 7, Appli
38	1260	98.9	323	15	US-10-389-640-13	Sequence 13, Appl
39	1260	98.9	324	14	US-10-314-861-16	Sequence 16, Appl
40	1260	98.9	345	14	US-10-338-411-5	Sequence 5, Appli
41	1260	98.9	345	15	US-10-389-640-5	Sequence 5, Appli
42	1260	98.9	346	14	US-10-338-411-9	Sequence 9, Appli
43	1260	98.9	346	15	US-10-389-640-9	Sequence 9, Appli
44	1260	98.9	359	14	US-10-033-717-33	Sequence 33, Appl
45	1260	98.9	359	14	US-10-033-717-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1

US-09-887-784-4
; Sequence 4, Application US/09887784
; Patent No. US20020177189A1

GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al

; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P

; CURRENT APPLICATION NUMBER: US/09/887,784

; CURRENT FILING DATE: 2001-06-19

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 4

; LENGTH: 239

; TYPE: PRT

; ORGANISM: Aequoria Victoria

US-09-887-784-4

Query Match 99.8%; Score 1271; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.2e-124;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MWSKGELFTGVVPIVELDGVNGHKFSVSGEGDATYKLT	1	MWSKGELFTGVVPIVELDGVNGHKFSVSGEGDATYKLT	Sequence 4, Appli
Db	1	MWSKGELFTGVVPIVELDGVNGHKFSVSGEGDATYKLT	1	MWSKGELFTGVVPIVELDGVNGHKFSVSGEGDATYKLT	Sequence 2, Appli
Qy	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSA	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSA	Sequence 30, Appl
Db	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSA	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSA	Sequence 32, Appl
Qy	121	VNRIELKIDFDKEDGNILGHKLEYNSHN	121	VNRIELKIDFDKEDGNILGHKLEYNSHN	Sequence 2, Appli
Db	121	VNRIELKIDFDKEDGNILGHKLEYNSHN	121	VNRIELKIDFDKEDGNILGHKLEYNSHN	Sequence 4, Appli
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS	Sequence 2, Appli
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS	Sequence 4, Appli

```
RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: US/10/296,953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match          99.8%; Score 1271; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.2e-124;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Qy 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 239

RESULT 3
US-10-270-223-6
; Sequence 6, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: Bioimage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; FILE REFERENCE: 3759-0126P
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match          99.8%; Score 1271; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 4.1e-124;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Qy 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.8%; Score 1271; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 1.5e-123;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 655 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 714
Qy 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 775 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 239
Db 835 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Qy 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.8%; Score 1271; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 1.5e-123;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 655 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 714
Qy 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 775 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 239
Db 835 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
```

```

; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match
  Best Local Similarity 99.8%; Score 1271; DB 14; Length 1132;
  Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPT 953

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-887-784-2

Query Match
  Best Local Similarity 99.1%; Score 1263; DB 9; Length 239;
  Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US2004007295A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, IEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match
  Best Local Similarity 99.2%; Score 1263; DB 12; Length 239;
  Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match
  Best Local Similarity 98.9%; Score 1260; DB 9; Length 239;
  Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPT 60
```

QY 61 LVTTLTSGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAGITLGMDELYK 239
DB 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAGITLGMDELYK 239
RESULT 9
US-09-999-745-4
; Sequence 4, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-999-745-4

Query Match 98.9%; Score 1260; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAGITLGMDELYK 239
DB 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAGITLGMDELYK 239

RESULT 10
US-09-866-538-4
; Sequence 4, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT

; ORGANISM: Aequorea victoria
US-09-866-538-4
Query Match 98.9%; Score 1260; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAGITLGMDELYK 239
DB 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAGITLGMDELYK 239
RESULT 11
US-09-797-496B-2
; Sequence 2, Application US/09797496B
; Publication No. US20030049597A1
; GENERAL INFORMATION:
; APPLICANT: Simon, Sanford M.
; APPLICANT: Chen, Yu
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof
; FILE REFERENCE: 600-1-267
; CURRENT APPLICATION NUMBER: US/09/797,496B
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described
US-09-797-496B-2

Query Match 98.9%; Score 1260; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAGITLGMDELYK 239
DB 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAGITLGMDELYK 239
RESULT 12
US-09-794-308-4
; Sequence 4, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA

APPLICANT: TSUEN, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REG1530
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-794-308-4

Query Match 98.9%; Score 1260; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQDFFKFSAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQDFFKFSAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 13
US-09-865-291-4
Sequence 4, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSUEN, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REG1550
CURRENT APPLICATION NUMBER: US/09/865,291
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-865-291-4

Query Match 98.9%; Score 1260; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQDFFKFSAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQDFFKFSAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 14
US-10-457-982-3
Sequence 3, Application US/10457982
Publication No. US2003021265A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Atsushi
APPLICANT: Liopis, Juan
APPLICANT: Wachter, Rebekka M.
APPLICANT: Remington, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/10/457,982
CURRENT FILING DATE: 2003-06-09
PRIOR APPLICATION NUMBER: US/09/602,641
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/172,063
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 239
TYPE: PRT
ORGANISM: Aequorea victoria
FEATURE:
NAME/KEY: VARIANT
LOCATION: (0)...(0)
OTHER INFORMATION: EGFP
US-10-457-982-3

Query Match 98.9%; Score 1260; DB 12; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQDFFKFSAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQDFFKFSAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 15
US-10-121-258-13
Sequence 13, Application US/10121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP1
CURRENT APPLICATION NUMBER: US/10/121,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24

Query Match 98.9%; Score 1260; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQDFFKFSAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQDFFKFSAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.9%; Score 1260; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKLTTLKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKLTTLKFICTTGKLPVPWPT 60
Qy 61 LVTTLSYGVCQFGRYPDPHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLYGVQCQFGRYPDPHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
Qy 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
Db 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:29
Job time : 35.7778 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds
(without alignments)
2224.817 Million cell updates/sec

Title: US-09-887-784-222V
Perfect score: 1274
Sequence: 1 MYSKGELFTGVVPIILVELD.....VLGVFTAAAGITLGMDELYK 239
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1238	97.2	238	1 JQ1514	green-fluorescent
2	104	8.2	785	2 H72228	hypothetical prote
3	91.5	7.2	861	2 H64102	leucine-trRNA ligas
4	90	7.1	632	2 T06586	DNA-binding protei
5	89.5	7.0	797	2 JC4078	protective surfac
6	89.5	7.0	808	2 F64102	protective surfac
7	89.5	7.0	887	2 E82590	leucyl-trRNA synthe
8	88.5	6.9	655	2 D83917	DNA topoisomerase
9	88.5	6.9	941	2 S29043	cellulase (EC 3.2.
10	88	6.9	578	1 I40794	dihydrolipoamide d
11	87.5	6.9	370	2 E70390	iron-sulfur cofact
12	87.5	6.9	860	2 AC0582	leucyl-trRNA synthe
13	87.5	6.9	2232	1 A36028	DNA-directed DNA p
14	87.5	6.9	2573	2 D71614	hypothetical prote
15	86	6.8	357	2 G81355	tRNA (uracil-5)-m
16	85.5	6.7	613	2 A99552	oligoendopeptidase
17	85.5	6.7	788	1 JDVLLH	DNA-directed DNA p
18	85.5	6.7	889	2 JC5576	inter-alpha-typsin
19	85	6.7	281	2 AD2052	hypothetical prote
20	84.5	6.6	425	2 C97354	hypothetical prote
21	84	6.6	353	2 E84941	imidazoleglycerol-
22	84	6.6	874	2 JC4930	S-layer protein pr
23	83.5	6.6	836	1 JDVLD	DNA-directed DNA p
24	83.5	6.6	1134	1 A60234	IgA Fc receptor pr
25	83.5	6.6	1164	1 FCSOAG	IgA Fc receptor pr
26	83	6.5	461	2 T06936	photosystem II chl
27	82.5	6.5	263	2 S53488	water-stress-induc
28	82.5	6.5	534	1 NICLMA	nitrogenase (EC 1.
29	82.5	6.5	740	2 G95153	neuraminidase, pro

ALIGNMENTS

RESULT 1

JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C:Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:1347277
A:Accession: JS0692
A:Molecule type: DNA
A:Residues: 1-107, 'S', 109-238 <PEA1>
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99, 'F', 'L', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 'O', 26-156, 'P', 158-171, 'K', 173-238 <INO>
A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384
R:Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009
A:Experimental source: clone gfp1
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24, 'O', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 209-238 <PRA5>
A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011
A:Experimental source: clone gfp2
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65692; PDB:1GFL
A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-95
A:Notes: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58953; MUID:98294543; PMID:9631087

hypothetical prote
synaptogamin o-p65
hypothetical prote
hypothetical prote
carboxy-terminal p
hypothetical prote
hypothetical prote
proprotein convert
alpha-amyase homo
lipoxigenase (EC 1
synergismenotropi
hypothetical prote
hypothetical prote
hypothetical prote
conserved hypothet
leucine-trRNA ligas


```
QY 200 -HYLSTOSALSADPNKR 216
Db 483 VAFSGTGSPKTPDPEWRK 500

RESULT 8
D83917
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: D83917
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai, T.
Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: D83917
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <STO>
A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05859.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2140
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 6.9%; Score 88.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK--PVSVEGEGDAT---YGKLTLEI-----CTTGKLPVWP 59
Db 63 NVTIHKQSVSDRGEMTGMHKLKPTPEVILTVLHAGGFGGGYATSGLHGVA 122

QY 60 TLVTTLSYGVCFSRYPDPMKHQDFKSAMPEGYVQER-----TIFPKDGG----- 105
Db 123 SVNALSEWLIVKIKRDGWVEQRFENGKGPSTLEKKGKTRGTGTHFKPDPTVFSTT 182

QY 106 --NYKTRAEVKFGDITLVNRIELKGDIFKEDGNILGHLEYNYNHNVIMADK----- 157
Db 183 NFNVETLSERLRAAFLKGLKILVDLDDTKEVPH-YEDGKAFVYELNEDKETLHPV 241

QY 158 -----QKNGIKVNFKRHNIEDGSVOLADHYQNTPIGDPVLLPDNHYLSTOSALSADPN 212
Db 242 VFNFGSNGIEIEFAQFN--DGYTENVLSFVNVRITKG-----GTHELCAKTAMTRAV 294

QY 213 NE 214
Db 295 NE 296

RESULT 9
S29043
cellulase (EC 3.2.1.4) - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C>Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
C:Accession: S29043; PC4404
R:Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
J. Gen. Microbiol. 136, 1327-1334, 1990
A>Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from
A:Reference number: S29043; MUID:91037937; PMID:2230718
A:Accession: S29043
A:Molecule type: DNA
A:Residues: 1-941 <OZA>
A:Cross-references: EMBL:M27420; NID:g142664; PIDN:AAA22304.1; PID:g142665
R:Shirai, T.; Yamane, T.; Hidaka, T.; Kuyama, K.; Suzuki, A.; Ashida, T.; Ozaki, K.; Ito, S.
J. Biochem. 122, 683-685, 1997
A>Title: Crystallization and preliminary X-ray analysis of a truncated family A alkaline
A:Reference number: PC4404; MUID:98060488; PMID:9399567
A:Accession: PC4404
A:Molecule type: protein
A:Residues: 228-584 <SHI>
A:Experimental source: strain KSM-635
```

```
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Superfamily: Bacillus sp. KSM-635 alkaline cellulase; S-layer repeat homology; Thermotoc
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:41-95/Domain: S-layer repeat homology <SLR1>
F:101-153/Domain: S-layer repeat homology <SLR2>
F:164-219/Domain: S-layer repeat homology <SLR3>
F:766-908/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match 6.9%; Score 88.5; DB 2; Length 941;
Best Local Similarity 20.7%; Pred. No. 28;
Matches 50; Conservative 32; Mismatches 62; Indels 97; Gaps 11;

QY 16 LVELDGVNKHKFSVSGEGEGDATYVKLTLCITGKLPVWPPLTVTLTSLGVGVCFSRY 75
Db 240 LVELNG-----QUTLAGE---DGT-----PVQLRGWMT---HGLQWFG-- 271

QY 76 PDHMKQHDFFKSAMPEGYVQERTIFPKDGNKYKTRAEVKFEGDTLVNRIELKGDIFKEDG 135
Db 272 -EIVNENAFVALSDWGSNMIRLAMYIGENGATNPEVK---DLVVEGIELA----- 319

QY 136 NILGHKLEYNYNHNVYIMADKQKNGIKVNFIRHNIEDGSV-----QLADHYQOQNTPIGD 191
Db 320 -----FEHDMYIVDWH---VHAPGDPADVYSGAYDFFEEIADHYKDH----- 360

QY 192 GPVLLPDNHYLSTOSALSADPN-----EKRDHMLV 222
Db 361 -----PKNHYIITWELANEPSPNNNGPGLTNDEKGWEAVKEYAEPIVEMLRKGDNMILV 415

QY 223 G 223
Db 416 G 416

RESULT 10
I40794
dihydrolipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum
N:Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase
hydrogenase complex chain E3; S-complex 50K chain
C:Species: Clostridium magnum
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
A:Accession: I40794
R:Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuechel, A.
J. Bacteriol. 176, 3614-3630, 1994
A>Title: Biochemical and molecular characterization of the Clostridium magnum acetoin de
A:Reference number: I40789; MUID:94266715; PMID:8206840
A:Accession: I40794
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-578 <KRU>
A:Cross-references: GB:L31844; NID:9472324; PIDN:AAA21748.1; PID:9472330
C:Function:
A:Description: catalyzes the oxidation of dihydrolipoamide to lipoamide using NAD
A:Pathway: acetoin dehydrogenase enzyme system
C:Superfamily: Alkaligenes dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase
C:Keywords: FAD; flavoprotein; lipoamide; NAD; oxidoreductase; redox-active disulfide
F:5-77/Domain: lipoyl/biotin-binding homology <LBP>
F:117-145/Region: beta-alpha-beta FAD nucleotide-binding fold
F:119-561/Domain: dihydrolipoamide dehydrogenase homology <DLID>
F:287-315/Region: beta-alpha-beta NAD nucleotide-binding fold
F:153-158/Disulfide bonds: redox-active #status predicted

Query Match 6.9%; Score 88; DB 1; Length 578;
Best Local Similarity 22.9%; Pred. No. 16;
Matches 57; Conservative 43; Mismatches 91; Indels 58; Gaps 13;

QY 10 TGWVPIIVELDGVNKHKFSVSGEGEGDATYVKLTLC-----FICTTGKLPVWPPLTVT 63
Db 255 TGSMPFIPPIEIR---GNKLS----GVIDST-GALSLESNPESIAIGGVIGVEFASIFN 305

QY 64 TLSYGVQCFSRYPDPMKHQHDFFKSAMPEGYVQERTIFPKDGNKYKTRAEVKFEGDTLVNRI 123
```

Db 306 SLGCKVSIIEMLPHILPPMDREISEI-----AKAKLIRDGININN 346

QY 124 IELKGDIFKEDG---NILGHKLEYNYNHNVIMADKQKN--GIKVNFKIRNIHEDGSVQ 178

Db 347 CKVTRIEQEGDGLKVSFIGDKGESIDVEKVLIAVGRRSNIEGLDVE-KIGVKTEGGSII 405

QY 179 LADHYQONT-----PTGD--GPVLLPDNHYLSQTQSALSKD-----PNEKRDMHVLVGFV--- 225

Db 406 VNDKMETNVEGIYAIGDCTGKIMLA--HVASDQGVVAENIMQNKMDYKTPACVYTK 463

QY 226 ---TAAGIT 231

Db 464 PELASVGLT 472

RESULT 11

E70390

iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus

N:Contains: L-cysteine sulfurtransferase (EC 2.8.1.1.-)

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999

C:Accession: E70390

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70390

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-370 <AQF>

A:Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: nifs1

C:Superfamily: nitrogen fixation protein nifs

C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase

F195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match 6.9%; Score 87.5; DB 2; Length 370;

Best Local Similarity 25.4%; Pred. No. 9.9;

Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY 4 KGEELFTGVV---PILVELD---GDVNGHFF-SVSGEG---EGDATYVGLTKLFICT 50

Db 164 KGVLPLTDAVQAIKPIELKNIYSATFSGHKFHAIKGSGFLYISDEANYEFLIVCGGQE 223

QY 51 TGKLP-----VPWPTLVTTLSYGVQCFSRYPDHMKQ-HDFFKSAMPEGYVQERTIFFKDD 104

Db 224 NGRSGTENNVVGLSLAKALEIIVSNFSRYQBLKRLDLFENLLLEA-LPDAQIVGKDA 282

QY 105 GNYKTRAEV---KFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNG 161

Db 283 ERSPSTSSVIMPKFFGAIEIVNKLSEKGYICSTGSACLSGEYEPNKMHLKMGFSQEKALRM 342

QY 162 IKVNFKIRENIED 174

Db 343 VRPSFGLLNKEE 355

RESULT 12

AC0582

leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S

C:Species: Salmonella enterica subsp. enterica serovar typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AC0582

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AC0582

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-860 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176

C:Genetics:

A:Gene: STY0699

C:Superfamily: leucine-tRNA ligase

Query Match 6.9%; Score 87.5; DB 2; Length 860;

Best Local Similarity 23.3%; Pred. No. 30;

Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLVPWPPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109

Db 314 TGEETPV-WAANFVLMMEYGTGAVMAVPGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370

QY 110 RAEVKEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIR 169

Db 371 SEQALTERKGLVFNSEGFDCGLAFEAFAFNAIADKL-----AEKGVGERKVNRYLR 418

QY 170 H-----NIEDGSVQLADHYQONTPIGDGPVLLPDNHYL-STQSALS 212

Db 419 DWGVSQRQYWGAPIPMVLTLEDGT-----LPTPDQLPVILPEVDVMDGITSPIKADP 471

RESULT 13

A36028

DNA-directed DNA polymerase (EC 2.7.7.7) II catalytic chain - yeast (Saccharomyces cerevi

N:Alternate names: DNA-directed DNA polymerase II chain A; protein N0825; protein YNL262v

C:Species: Saccharomyces cerevisiae

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: A36028; B36028; S60919; S63235; S65121

R:Morrison, A.; Araki, H.; Clark, A.B.; Hamatake, R.K.; Sugino, A.

Cell 62, 1143-1151, 1990

A:Title: A third essential DNA polymerase in Saccharomyces cerevisiae.

A:Reference number: A36028; MUID:90381771; PMID:2169349

A:Accession: A36028

A:Molecule type: DNA

A:Residues: 1-2222 <MOR>

A:Cross-references: GB:IM60416; GB:M36724; NID:gl71408; PIDN:AAA88711.1; PID:gl71409

A:Accession: B36028

A:Molecule type: protein

A:Residues: 1214-1216,'X',1218-1221 <MO2>

R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.M.

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV fr

A:Reference number: S60909

A:Accession: S60919

A:Molecule type: DNA

A:Residues: 1-2221 <SEN>

A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247

R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S63235

A:Accession: S63235

A:Molecule type: DNA

A:Residues: 1-2222 <SEN>

A:Cross-references: EMBL:271538; NID:g1302316; PIDN:CAA96169.1; PID:g1302317; GSPDB:GN000

A:Experimental source: strain S288C

R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.

Yeast 12, 505-514, 1996

A:Title: The sequence of a 24 152 bp segment from the left arm of chromosome XIV from Sac

A:Reference number: S65111; MUID:96310631; PMID:8740425

A:Accession: S65121

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2221 <SEP>

A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds
(without alignments)
1931.085 Million cell updates/sec

Title: US-09-887-784-222V
Perfect score: 1274
Sequence: 1 MVSGBELFTGVVPLVELD.....VLGVFTAAIGTGMDELYK 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1244	97.6	238	1 GFP_AEQVI	P42212 aequorea vi
2	91.5	7.2	861	1 SYL_HAEIN	P43827 haemophilus
3	89.5	7.0	795	1 D152_HAEIN	P44935 haemophilus
4	89.5	7.0	797	1 D151_HAEIN	P46024 haemophilus
5	89.5	7.0	879	1 SYL_XYLFA	Q9pbg8 xylella fas
6	88.5	6.9	793	1 D153_HAEIN	Q32629 haemophilus
7	88.5	6.9	941	1 GUN_EACS6	P19424 bacillus sp
8	87.5	6.9	860	1 SYL_SALTY	Q8z8h5 salmonella
9	87.5	6.9	860	1 SYL_SALTY	Q8zqz6 salmonella
10	87.5	6.9	2222	1 DPOE_YEAST	P21951 saccharomyc
11	87	6.8	689	1 AC2L_HUMAN	Q9nub1 homo sapien
12	86.5	6.8	533	1 CP51_CANGA	P50859 candida gla
13	86.5	6.8	879	1 SYL_XYLFT	Q87c65 xylella fas
14	86.5	6.8	1603	1 VIT4_CABEL	P18947 caenorhabdi
15	86	6.8	357	1 TRMA_CAMJE	Q9pb92 campylobact
16	85.5	6.7	613	1 PEPE_MYCPU	Q98dp0 mycoplasma
17	85.5	6.7	788	1 DPHL_HBHEA	P13846 heron hepat
18	85.5	6.7	886	1 ITH3_MESAU	P97280 mesocricetu
19	85	6.7	504	1 YC03_KLEPN	Q48449 klebsiella
20	84.5	6.6	501	1 AMPA_WIGBR	Q8d295 wigglewort
21	84.5	6.6	859	1 SYL_SHON	Q8ehp4 shewanella
22	84	6.6	353	1 HIS7_BUCAI	P57203 buchnera ap
23	84	6.6	366	1 SET7_HUMAN	Q8wt86 homo sapien
24	84	6.6	874	1 SLAP_BACLI	P49052 bacillus li
25	83.5	6.6	538	1 GRBE_RAT	Q88900 rattus norv
26	83.5	6.6	1164	1 BAG_STRAG	P27951 streptococc
27	83	6.5	461	1 PSBC_CVAPA	P48104 cyanophora
28	82.5	6.5	533	1 NIPD_CLOPA	P00467 clostridium
29	82	6.4	439	1 PY62_DLSOM	P24506 discopyge o
30	82	6.4	682	1 SRC_ECOLI	P23865 escherichia
31	82	6.4	752	1 NECI_RAT	P28840 rattus norv
32	82	6.4	774	1 AMY2_SCHPO	O42918 schizosacch
33	81.5	6.4	589	1 SYD_HAEIU	Q7vnf0 haemophilus

RESULT 1				
ID	GFP_AEQVI	STANDARD;	PRT;	238 AA.
AC	P42212; Q17104; Q27903;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent			
RT	protein.";			
RL	Gene 111:229-233(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and			
RT	fluorescence characteristics of the recombinant protein.";			
RL	FEBS Lett. 341:277-280(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97299832; PubMed=9154981;			
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;			
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent			
RT	protein by modification of its codon usage.";			
RL	Plant Mol. Biol. 33:989-999(1997).			
RN	[4]			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea			
RT	green-fluorescent protein.";			
RL	Biochemistry 32:1212-1218(1993).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormeo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,			
RA	Remington S.J.;			
RT	"Crystal structure of the Aequorea victoria green fluorescent			
RT	protein.";			
RL	Science 273:1392-1395(1996).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein.";			
RL	Nat. Biotechnol. 14:1246-1251(1996).			

Q58743 methanococc
Q99nbl mus musculu
P91679 drosophila
Q25443 helicobacte
O00116 homo sapien
Q8xbn8 escherichia
Q8fjy9 escherichia
P07813 escherichia
Q08694 chlamydomon
Q9pk7 chlamydia m
P94136 azorhizobiu
P36924 bacillus ce

ALIGNMENTS

[7]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
RX MEDLINE=98455509; PubMed=9782051;
RA Wachter R.M., Eislinger M.A., Kallio K., Hanson G.T., Remington S.J.;
RT "Structural basis of spectral shifts in the yellow-emission variants
of green fluorescent protein."; Structure 6:1267-1277(1998).
RL [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=99238303; PubMed=10220315;
RA Eislinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
RT "Structural and spectral response of green fluorescent protein
variants to changes in pH."; Biochemistry 38:5296-5301(1999).
RL [9]
CC -1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the
blue chemiluminescence of the protein aequorin into green
fluorescent light by energy transfer. Fluoresces in vivo upon
receiving energy from the Ca(2+)-activated photoprotein aequorin.
Absorbs light maximally at 395 nm and exhibits a smaller
absorbance peak at 470 nm. The fluorescence emission spectrum
peaks at 509 nm with a shoulder at 540 nm.
CC -1- SUBUNIT: Monomer.
CC -1- TISSUE SPECIFICITY: Photocytes.
CC -1- PTM: Contains a covalently attached chromophore, which is composed
of modified amino acid residues. The chromophore is formed upon
cyclization of the residues Ser-dehydroTy-Gly.
CC -1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
chimeric proteins of GFP linked to other proteins where it
functions as a fluorescent protein tag. GFP tolerates N- and C-
terminal fusion to a broad variety of proteins. It has been
expressed in bacteria, yeast, slime mold, plants, Drosophila,
zebrafish, and in mammalian cells. As a noninvasive fluorescent
marker in living cells, it allows for a wide range of applications
where it may function as a cell lineage tracer, reporter of gene
expression, or as a measure of protein-protein interactions.
CC -1- DATABASE: NAME=Protein Spotlight;
NOTE=Issue 11 of June 2001;
WWW="http://www.expasy.org/spotlight/articles/sptl011.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M62654; AAA27722.1; -;
DR EMBL; M62653; AAA27721.1; -;
DR EMBL; L29345; AAA58246.1; -;
DR EMBL; X96418; CAA65278.1; -;
DR PIR; JS0692; JQ1514.
DR PDB; 1B9C; 17-NOV-00.
DR PDB; 1BFP; 07-JUL-97.
DR PDB; 1C4P; 14-JUN-00.
DR PDB; 1EMA; 08-NOV-96.
DR PDB; 1EMB; 16-JUN-97.
DR PDB; 1EMC; 20-AUG-97.
DR PDB; 1EMD; 20-AUG-97.
DR PDB; 1EMF; 20-AUG-97.
DR PDB; 1EMG; 12-MAY-99.
DR PDB; 1EMK; 20-AUG-97.
DR PDB; 1EML; 20-AUG-97.
DR PDB; 1EMW; 20-AUG-97.
DR PDB; 1F09; 17-NOV-00.
DR PDB; 1F0B; 17-NOV-00.
DR PDB; 1GFL; 11-JAN-97.
DR PDB; 1HCJ; 15-JAN-92.
DR PDB; 1HUY; 04-JUL-01.
DR PDB; 1JBZ; 07-JAN-03.
DR PDB; 1KPS; 28-AUG-02.
DR PDB; 1KVP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.
DR PDB; 1KYS; 10-APR-02.
DR PDB; 1YFP; 28-OCT-98.
DR PDB; 2EMD; 20-AUG-97.
DR PDB; 2EMN; 20-AUG-97.
DR PDB; 2EMO; 20-AUG-97.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
KW Luminescence; 3D-structure.
FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).
FT MOD_RES 66 66 2,3-DIDEHYDROTYROSINE.
FT VARIANT 100 100 F -> Y.
FT VARIANT 108 108 T -> S.
FT VARIANT 141 141 L -> M.
FT VARIANT 219 219 V -> I.
FT CONFLICT 2 2 S -> G (IN REF. 3).
FT CONFLICT 25 25 H -> Q (IN REF. 2).
FT CONFLICT 80 80 Q -> R (IN REF. 3).
FT CONFLICT 157 157 Q -> P (IN REF. 2).
FT CONFLICT 172 172 E -> K (IN REF. 2).
FT HELIX 4 8
FT STRAND 12 22
FT TURN 23 24
FT STRAND 25 36
FT TURN 37 40
FT STRAND 41 48
FT TURN 49 50
FT HELIX 57 60
FT TURN 61 63
FT HELIX 69 71
FT STRAND 73 73
FT HELIX 76 81
FT HELIX 83 86
FT TURN 87 90
FT STRAND 92 100
FT TURN 101 102
FT STRAND 105 115
FT TURN 116 117
FT STRAND 118 128
FT TURN 132 133
FT TURN 135 139
FT STRAND 141 141
FT STRAND 148 155
FT TURN 156 159
FT STRAND 160 171
FT TURN 172 173
FT STRAND 176 187
FT STRAND 199 208
FT TURN 211 212
FT STRAND 217 227
SQ SEQUENCE 238 AA; 26886 MW; EASA6F21FBFB6E05 CRC64;
Query Match 97.6%; Score 1244; DB 1; Length 238;
Best Local Similarity 97.9%; Pred. No. 3 1e-95;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 2 VSKGEELFTGVWPILVELDGVNNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVWPILVELDGVNNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVWPPTL 60
QY 62 VTTLVSVQCFSSRYPDHMKQHDFFKSAMPEGVYQERTIPKDDGNKYKTRAEVKEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTFSYGVQCFSSRYPDHMKQHDFFKSAMPEGVYQERTIPKDDGNKYKTRAEVKEGDTLV 120
QY 122 NRTELKGDIFKEDGNILGHKLEYNYNHNVYIMADQKNGIKVNFKIRHNIEDGSVOLAD 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRTELKGDIFKEDGNILGHKLEYNYNHNVYIMADQKNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVGVFTAGITLGHDELYK 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```
Db 181 HYQNTPTIGDGPVLLPDNHYLSQSALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
RESULT 2
SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32774; AAC22581.1; -.
CC PIR; H64102; H64102.
CC TIGR; HI0921; -.
CC HAMAP; MF 00049; -.
CC InterPro; IPR002302; Leu-tRNA-synt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_fiers_edit.
CC Pfam; PF00133; tRNA-synt 1; 1.
CC PRINTS; P00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FE7 CRC64;
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 6.9;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVPPTLVTLLSYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQVERIFFKX-----103
Db 314 TGDKLPV-IVANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKYSLPKQVIAPLA 364
```

```
QY 104 DGNKYTRAEVKFGDGLVNRILKGLIDFKEDGNILGHKLEYNYNHNVYIMADK-QKNGI 162
Db 365 DEEIDLTKQAFVHGKLVNSDEDFGKNF--DGAENG-----IADKLEKLG 408
QY 163 ---KVNFKIRH-----NIEDGSVQLADHYQQTPTIGDGPVLLPDNHYL- 202
Db 409 GKRQVNYRLRDWGVSRQRYWGAPIPMLTLLENGDVVPA-----PMEDLPILPEDVYMD 461
QY 203 STQSALSQDPN 213
Db 462 GVKSPINADPN 472
RESULT 3
D152_HAEIN STANDARD; PRT; 795 AA.
AC F44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN HI0917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32773; AAC22575.1; -.
CC TIGR; HI0917; -.
CC InterPro; IPR000184; Bac_surfag_D15.
CC Pfam; PF01103; Bac_surfaceAg; 1.
KW Antigen; Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 795 AA; 87478 MW; B85691FC228E6D44 CRC64;
Query Match 7.0%; Score 89.5; DB 1; Length 795;
Best Local Similarity 22.4%; Pred. No. 9.2;
Matches 49; Conservative 29; Mismatches 78; Indels 63; Gaps 11;
QY 65 LSYGVQCFSRYPDHMKQHDFF-----FKSAMPEGYVQE-----RTI 99
Db 427 IGYGESGISYQASVKQDNFLGTGAASVIACTKNDYGTSLNLTGTFPYFTKDGVS LGNV 486
QY 100 FKDDGKNYKTRAEVKFEGDTLVNRILKGLIDFKEDGNI---LGH-----KLEYNYS 148
```



```
Db 487 FFENYDMSKSDTSSNYKRTTYGNSVTL-GFPVNNENSYVGLGHTYNNKISNFALEYN--- 542
Qy 149 HNVYIMADKQK-NGIKVNFKIRHNIEDSGVQLADHYQQ-----NTPIGDGPVLL 196
Db 543 RNLXIQSMKFKNGIKTN-----DFDFSGWNYSNLRGYPFTKGVKASLG-GRVTI 593
Qy 197 P--DNHYLSTQSALS KDPNEKRDHVMVLGVFTAAAGITLG 233
Db 594 PGSDNKYYKLSADVQGFYPLDRDLHLWVVSASAGSAGYANG 632

RESULT 4
ID D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) Outer membrane protein D15.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RT Haemophilus influenzae.";
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Minna / Serotype B, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U13961; AAA85645.1; -
CC DR EMBL; U60832; AAB61974.1; -
CC DR EMBL; U60833; AAB61976.1; -
CC DR F01; JC4078; JC4078.
CC DR InterPro; IPR000184; Bac_surfAg_D15.
CC DR Pfam; PF01103; Bac_surface_Ag; I.
CC KW Antigen; Outer membrane; Signal.
CC FT SIGNAL 1 19 POTENTIAL.
CC FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
CC FT SEQUENCE 797 AA; 87675 MW; 2F93DE53869AF1B CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 797;
Best Local Similarity 22.4%; Pred. No. 9.2;
Matches 49; Conservative 29; Mismatches 63; Gaps 11;

Qy 65 LSYGVQCFSPYDPMKQHDR-----FKSAMPEGYVOE-----RTI 99
Db 427 IGVGTSGISYQASVKQDNFLGTGAASVIACTKNDYGTSNLYGTPYFTKDGVSIGGNV 486
Qy 100 PFKDDGNKYTRAEVKFEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148
```

```
Db 487 FFENYDMSKSDTSSNYKRTTYGNSVTL-GFPVNNENSYVGLGHTYNNKISNFALEYN--- 542
Qy 149 HNVYIMADKQK-NGIKVNFKIRHNIEDSGVQLADHYQQ-----NTPIGDGPVLL 196
Db 543 RNLXIQSMKFKNGIKTN-----DFDFSGWNYSNLRGYPFTKGVKASLG-GRVTI 593
Qy 197 P--DNHYLSTQSALS KDPNEKRDHVMVLGVFTAAAGITLG 233
Db 594 PGSDNKYYKLSADVQGFYPLDRDLHLWVVSASAGSAGYANG 632

RESULT 5
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9PBG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5G; PubMed=10910347;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsumura A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004031; AAF84975.1; ALT_INIT.
CC HAMAP; MF_00049; -; 1.
CC InterPro; IPR002302; Leu-tRNA-synt1a.
CC
```



```
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; Valrsls_edit.
DR Pfam; PF00133; tRNA-synt_1.1.
DR PRINTS; PR00985; TRNASYNTHLSU.
DR TIGRFAMs; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "RMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99796 MW; 9FDCB992092919E CRC64;

Query Match
Best Local Similarity 22.2%; Pred. No. 10;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPMPTLVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFPKDQGNV-- 107
DB 321 TNEQLPV-VVANFVLMAYGTGAVMAVPGHQRQDEP--ANKYGLPIRQVIALKEPKNOD 377
QY 108 -----KTRAEVKEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVI 153
DB 378 STWEPDVRWDYADKTR---EPE---LINSAEFDGLDYOQDAFEVLAERPE----- 421
QY 154 MADKQKNG-IKYNFKIRHNIEDGSVQLADHYQQNTPI-----GDGPVLLPDN 199
DB 422 ---RQGRQRRVNRUR-----DWGVSQRVYGCPIVYICPGAVPVEDQLPVILPEN 474
QY 200 -HYLSTQSALSADPNKPR 216
DB 475 VAFSGTGSGPIKTDPEWRK 492

RESULT 6
D153.HAEIN
ID D153 HAEIN STANDARD; PRT; 793 AA.
AC O32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RX MEDLINE=97427952; PubMed=9284140;
RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U60834; AAB61977.1; -.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac surface Ag; 1.
KW Antigen; Outer membrane; Signal.
```

```
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A14 CRC64;

Query Match
Best Local Similarity 22.8%; Pred. No. 11;
Matches 50; Conservative 27; Mismatches 79; Indels 63; Gaps 11;

QY 65 LSYGVQCFSRYPDHMKQHDFF-----FKSAMPEGYVOE-----RTI 99
DB 427 IGYGTESGISYQTSIKQDNFLGTGAAGVSTAGTKNDYGTSVNLGYTEPYFTKDGVSIGCNI 486
QY 100 FFKDDGNKYKTRAEVKEFGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148
DB 487 FFENYDNSKSDTSSNYKRTTYGTSNVTL-GFPVNNNSYVVGHTYKNSFALEYN--- 542
QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGSVQLADHYQQ-----NTPFGDGPVLL 196
DB 543 RNLYIQSMKFKNGIKTN-----DFDPSFGWYNSLNRYFTPKGVKASLG-GRVTI 593
QY 197 P--DNHYLSTQSALSADPNKPRDHVVLGVFVTAAGITLG 233
DB 594 PGSDNKYYKLSADVQGFPLDRDRHWVSASAGSYANG 632

RESULT 7
GUN_BACS6
ID GUN_BACS6 STANDARD; PRT; 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Alkaline cellulase).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=2230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT "Molecular cloning and nucleotide sequence of a gene for alkaline
RT cellulase from Bacillus sp. KSM-635."
RL J. Gen. Microbiol. 136:1327-1334(1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27420; AAA22304.1; -.
DR PIR; S29043; S29043.
DR PDB; 1G01; 31-DEC-02.
DR PDB; 1G0C; 31-DEC-02.
DR InterPro; IPR005086; CBM_17_28.
DR InterPro; IPR008979; Gal_bind like.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR001119; SLH.
DR Pfam; PF03424; CBM_17_28; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS01072; SLH DOMAIN; 2.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat;
KW 3D-structure.
```

```

FT SIGNAL 1 29
FT CHAIN 30 941
FT DOMAIN 40 99
FT DOMAIN 100 151
FT DOMAIN 152 225
FT ACT SITE 373 373
FT ACT SITE 485 485
SQ SEQUENCE 941 AA; 104628 MW; BEA2AC3BB169BFADA CRC64;

Query Match 6.9%; Score 88.5; DB 1; Length 941;
Best Local Similarity 20.7%; Pred. No. 14;
Matches 50; Conservative 32; Mismatches 62; Indels 97; Gaps 11;

QY 16 LVLELDGVNGHKFSVSGEGBDATYKGLTKLKFCTTGKLPVWPPTLVTLTSLYGVQCFSRV 75
DB 240 LVELNG-----QUTLAGE---DGT-----PVQRGMST---HGLQWFG-- 271
QY 76 PDHMKQHDFFKSAHPGYSYQVQERTIFFKDCDGNKYTRAEVKEPFGDTLVNRIELKGDIDFKEDG 135
DB 272 -EIVNENAFVLSNDWGSNMIRLAMYIGENGYATNPEVK---DLVYEGIELA----- 319
QY 136 NILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSV-----QLADHYQOQNTPIGD 191
DB 320 -----FEHDMYVIIVDMH---VHAPGDPRAVDYSGAYDFBEIADHYKDH----- 360
QY 192 GPVLLPDNHYLSQTSALSADPN-----EKRDHMLV 222
DB 361 -----PNHYIIVELANEPSPNNGGFLTNDEKGWEAVKEYAEPIVEMLRKGDNMILV 415
QY 223 G 223
DB 416 G 416

RESULT 8
SYL_SALTY STANDARD; PRT; 860 AA.
AC Q8Z8H5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (SC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR STV0699 OR T2219.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gea P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).

```

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627267; CAD05125.1; -
CC EMBL; AB016841; AAC069822.1; -
CC HAMAP; MF_00049; -; 1.
CC InterPro; IPR002302; Leu-trNASyntla.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR01412; tRNA-synt_1.
CC InterPro; IPR009008; VALRS_1IERS_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRfam; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 860 AA; 96940 MW; 2F95E480BBAB23C4 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 15;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVPWPTLVTLTSLYGVQCFSRYPDHMKQHDFFKSAHPGYSYQVQERTIFFKDCDGNKYT 109
DB 314 TGEIEPV-WAANFVLMYGTGAVMAVPGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKPEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIR 169
DB 371 SQALTEKGVLFNSGFEFDGLAFEAFAFNATADKL-----AEKGVGERKVNRLR 418
QY 170 H-----NIEDGSVOLADHYQOQNTPIGDGPVLLPDNHYL-STQSALSKOP 212
DB 419 DWGVSQRQYWGAPIPMWLTLEDGTV-----LPTPEDQLPVILPDVMDGITSPKADP 471

RESULT 9
SYL_SALTY STANDARD; PRT; 860 AA.
AC Q8Z0Z6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (SC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR STM0648.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).

```

```

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; A5008725; AAL19599.1; -.
CC StyGene; SG????; leuS.
CC HAMAP; MF 00049; -. 1.
CC InterPro; IPR002302; Leu-TRNAsynt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_fiers_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA-TRNA-LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
FT SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;
Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 15;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;
QY 50 TTGKLPVPTLTVTLSYGVQCSRPYDHMKQHQDFPKSAMPEGYQVERIFFPKDGNKYK 109
DB 314 TGEIEPV-WAANFVLMYEGTGAVMAYPGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKEGDTLVNRIELKIDFKEDCNILGHKLEYNHNVYIMADKQNGIKVNFKR 169
DB 371 SEQALTEKGVLFNSGEFGDLAEAFNAIADKL-----AEKGVGERKVNRLR 418
QY 170 H-----NIEDGSVLADHYQQNTPIGDGVLPLPDNHYL-STQSALSKDP 212
DB 419 DWGVSQRVYGAPIPWVLTLEDGTV-----LPTPEDQLPVILPDVMDGITSPIKADP 471
RESULT 10
DPOE YEAST STANDARD; PRT; 2222 AA.
AC P21951;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
DE polymerase II subunit A).
GN POL2 OR DUN2 OR YNL262W OR N0825.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.
RX MEDLINE=90381771; PubMed=2169349;
RA Morrison A., Araki H., Clark A.B., Hamatake R.K., Sugino A.;
RT "A third essential DNA polymerase in S. cerevisiae.";
RL Cell 62:1143-1151(1990).
RN [2]
RP SEQUENCE OF 1-2221 FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2

```

```

RT Genes.";
RL Yeast 12:505-514 (1996).
RN [3]
RP TEMPERATURE SENSITIVE MUTANTS.
RX MEDLINE=92164663; PubMed=1537345;
RA Araki H., Kopp P.A., Johnson A.L., Johnston L.H., Morrison A.,
RA Sugino A.;
RT "DNA polymerase II, the probable homolog of mammalian DNA polymerase
RT epsilon, replicates chromosomal DNA in the yeast Saccharomyces
RT cerevisiae.";
RL EMBL J. 11:733-740 (1992).
CC -!- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
CC REPLICATION.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -!- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30
CC kDa, AND 29 kDa).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNITS B AND C.
CC -!- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
CC alpha, beta, gamma, delta, and epsilon which are responsible for
CC different reactions of DNA synthesis.
CC -!- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60416; AAA88711.1; -.
CC EMBL; X92494; CAA63235.1; -.
CC EMBL; Z71538; CAA96169.1; -.
CC PIR; A36028; A36028.
CC GenOnline; 143268; -.
CC SGD; S0005206; POL2.
CC GO; GO:0000731; P:DNA repair synthesis; IMP.
CC InterPro; IPR006172; DNA_pol_B.
CC InterPro; IPR006134; DNA_pol_B_dom.
CC InterPro; IPR006133; DNA_pol_B_exo.
CC Pfam; PF00136; DNA_pol_B_1.
CC Pfam; PF03104; DNA_pol_B_exo; 1.
CC SMART; SM00486; POLBC; 1.
CC PROSITE; PS00116; DNA POLYMERASE B; FALSE NEG.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Zinc-finger; Nuclear protein.
FT ZN FING 2108 2181 POTENTIAL.
FT VARIANT 644 644 M -> I (IN POL2-9 TS MUTANT).
FT VARIANT 710 710 P -> S (IN POL2-18 TS MUTANT).
SQ SEQUENCE 2222 AA; 255669 MW; CBCDDE2AB147D65B CRC64;
Query Match 6.9%; Score 87.5; DB 1; Length 2222;
Best Local Similarity 28.2%; Pred. No. 46;
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;
QY 54 LVPWMP-TLVTTLSYGVQCSFYDPHM-----KQHDFPKSAMPEGYV-----QERTI 99
DB 883 LPKSPETVFFFTLENGKKLYLSPCSMLNRYHQKFTNHQYQELKDPLNVIYETHSENTI 942
QY 100 FFKDDGNKYTR--AEVKFEGDTLVNR-----IELKGDIDFKEDGNILGHKLEYN 147
DB 943 PFEVDGPKAMILPSSKEEGKIKRYAVFNEDGSLAELKGFELKRGEL---QLKNFQ 999
QY 148 S--HNVIYIMAD 156
DB 1000 SDIFKVFLEGD 1010
RESULT 11

```



```
Db 473 FFGIVPVLMDKGSV-----VEGSNVSGALCIS-----QAWPGMARTI--- 510
QY 69 VQCFSRYPDHMKHQDFKFSAMPGEYQERTIFFKDGNYKTRA---EVKPEGDTLVNRIE 125
Db 511 -----YGDHQRFDVAFKAYP-GY-----YFTGDGAYTEGGYQITGRMDVI----- 553
QY 126 LKGIDPFKDNILGHKL 142
Db 554 -----NISGRL 560

RESULT 12
CP51_CANGA STANDARD; PRT; 533 AA.
ID CP51_CANGA
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIA1) (Sterol 14-
alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7969540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rosser M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIAL) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L40389; AAB02329.1; -.
CC EMBL; S75389; AAB32679.1; -.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; p450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
KW Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
```

```
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best local Similarity 21.4%; Pred. No. 10;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;

QY 25 GHKFSVS---GEGEGDATYKGLTKFKICTGKLPVDPWPTLVTTLSYGVQCFSRYPDH--M 79
Db 109 GHEFIENAKLADVSAEAAYSHL-----TTTFVGKGVYDCPNRLM 149
QY 80 KQHDFFSAM-PEGYV-----QERTIFFKDGNYKTRAEVKPEGDTLVNRIELKGIDF 131
Db 150 EQKKFVKGALTKEAFVRYVPLIAEIIKYFRNSKFNENNSGIVDMVMSQPEM--TIF 207
QY 132 KEDGNILGHKLEYNVSHNVIMADKQNGIKVKNFKIRHNIEDSGVQLADHYQNTPIGD 191
Db 208 TARSRLLGKEMRDKLDTPFAYLYSLDKGFTPINF-VFPNLPLEHYRKRDHAQAIS--- 263
QY 192 GPVLLPDNHYLSTQSALSQDPNEKRD 217
Db 264 -----GTYSMLIKERKND 278

RESULT 13
SYL_XYLFT STANDARD; PRT; 879 AA.
ID SYL_XYLFT
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR PDI230.
OS Xylella fastidiosa (strain Temecual / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorty H., Tsai S.M.,
RA Carter H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.B., Kimura E.F., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.B., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.B. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa";
RL J. Bacteriol. 185:1018-1026(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF012557; AAO29080.1; ALT_INIT.
DR HAMAP; MF_00049; -; 1.
```

```
DR InterPro; IPR002302; Leu-TRNAsyntla.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009008; ValRS_fiers_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PRO0985; TRNASYNTHLSU.
DR TIGRFAMS; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
FT SEQUENCE 879 AA; 99823 MW; 4C2EE0188FDC497E CRC64;
SQ
Query Match 6.8%; Score 86.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 18;
Matches 44; Conservative 28; Mismatches 69; Indels 57; Gaps 10;
QY 50 TTCKLPVPWPTLVTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107
DB 321 TNEQLPV-WVAVFLVWYGTGAVMVGPHGDQDEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 108 -----KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHVYI 153
DB 378 SIWEPDVRDWDYADKTR---EFE--LINSAPFDGLDYQGAPEVLAERFE----- 421
QY 154 MADKQKNG-IKVNFKIRHNIEDGSVQLADHYQNTPI-----GDGVPVLLPDN 199
DB 422 ---ROGRQRRVYRLR---DMGVSQRQYWGCFPIVYICPTGCAVPVPENQLPVILPEN 474
QY 200 -HYLSTQSALS KDPNEKR 216
DB 475 VAFSGTGSPKTPDFEWRK 492
RESULT 14
VIT4_CABEL STANDARD; PRT; 1603 AA.
AC FL8947; Q9BPP3;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
GN VIT-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T.; Spieth J.; Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=85269643; PubMed=4022780;
RA Spieth J.; Denison K.; Kirtland S.; Cane J.; Blumenthal T.;
RT "The C. elegans vitellogenin genes; short sequence repeats in the
RT promoter regions and homology to the vertebrate genes."
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -1- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
CC building the intestine of adult hermaphroditic individuals; they
CC are cotranslationally secreted into the body cavity and
CC subsequently taken up by the gonad.
CC -1- SIMILARITY: Contains 1 VWFD domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
CC EMBL; AC024137; AAK09074.1; --
CC EMBL; M11498; AAZ28163.1; --
CC EMBL; X02754; CAA26531.1; --
CC PIR; A43084; A43084.
CC WormPep; F59D8.2; CE26817.
CC InterPro; IPR001747; Lipid_transprt_N.
CC InterPro; IPR001846; VWFD.
CC Pfam; PF001347; Vitellogenin_N; 1.
CC SMART; SM00638; LPD_N; 1.
CC SMART; SM00216; VMD; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1603 VITELLOGENIN 4.
FT DOMAIN 1308 1455 VWFD.
FT CONFLICT 30 30 Y -> V (IN REF. 3).
FT CONFLICT 169 169 L -> V (IN REF. 2).
FT CONFLICT 183 187 EVAYT -> RSLRH (IN REF. 2).
FT CONFLICT 275 275 T -> S (IN REF. 2).
SQ SEQUENCE 1603 AA; 186307 MW; E303170325BC99BB CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 1603;
Best Local Similarity 23.4%; Pred. No. 37;
Matches 52; Conservative 32; Mismatches 69; Indels 69; Gaps 12;
QY 1 MYSKGEELFTGVPIILVELDGVNGHKPSVSGEGDATYGLTLKFICTTGKLPVWPWT 60
DB 162 MESDKDSLFFNVHEKTMGDCV---AYTIIVEG-GKTIYTKVNFDKCITR-----PE 211
QY 61 LVTTLISYGVCFSRYPDHMKQHDFFKSAMPEG-YVQERTIF---FKDDG----- 105
DB 212 TAYGLRFGSEC-----KECKEGQFVQPTVYVTFKNEKLQSEVNSIYT 257
QY 106 -----NYKTRAVKPEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHVIMAD 156
DB 258 LNVNGQEVVKSETRAKTVFVEESKINR-EIK-----KVGPKKEIYVSMENKLEIQ 308
QY 157 KQKNG-----IKVNFKIRHNIEDGSVOLADHYQNTP 188
DB 309 FYKQDKAEVNFKAIEISQKV-EQLEEIFRQIQEH-EQNTTP 348
RESULT 15
TRMA_CAMJE STANDARD; PRT; 357 AA.
AC Q9PP92;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5-)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-
DE methyltransferase) (RUMT).
GN TRMA OR CJ0831C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J.; Whittam B.W.; Mungall K.; Ketley J.M.; Churcher C.;
RA Basham D.; Chillingworth T.; Davies R.M.; Feltwell T.; Holtroyd S.;
RA Jagels K.; Karlyshev A.V.; Moule S.; Pallen M.J.; Penn C.W.;
RA Quail M.A.; Rajandream M.A.; Rutherford K.M.; van Vliet A.H.M.;
RA Whitehead S.; Barrell B.G.;
```

RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -!- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
CC 54 (M-5-U54) in all tRNA (By similarity).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing thymine.
CC -!- SIMILARITY: Belongs to the RNA M5U methyltransferase family. TrmA
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL139076; CAB73096.1; -.
DR PIR; G81355; G81355.
DR HAMAP; MF 01011; -; 1.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR001566; TrmA.
DR PROSITE; PS01230; TRMA_1; 1.
DR PROSITE; PS01231; TRMA_2; FALSE NEG.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY
FT ACT_SITE 315 315 BY SIMILARITY.
FT SEQUENCE 357 AA; 42276 MW; CEC5328347CEE497 CRC64;
SQ -----
Query Match 6.8%; Score 86; DB 1; Length 357;
Best Local Similarity 24.8%; Pred. No. 6.9;
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;
QY 80 KQHDFPKSAMPEGYQVQERTIFFKDDGNKYKTRAEVKF--EGDTLV-----NRIELKG 128
DB 14 EKHSFTKKYFKGFYTKDKLFASKDKHRTAELSPVHENDTLFYAMFDPKSKKKYIIEV 73
QY 129 IDFKED-----GNILGHKLEYNNSHNVIMADKQNGIKVNFKRHNIE 173
DB 74 LDFADEKICAFMPRLLEYLRQDNKLKEL-----EGVEFLATKQE--LSITLLYHKNIE 125
QY 174 D 174
DB 126 D 126

Search completed: June 21, 2004, 15:55:23
Job time : 6.55556 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1240	97.3	238	2	Q8GHE2	Q8ghe2 azotobacter
2	1237	97.1	238	5	Q93125	Q93125 aequeorea vi
3	1235	96.9	238	2	Q8GHE4	Q8ghe4 azomonas ag
4	1234	96.9	238	2	Q8GHE3	Q8ghe3 azotobacter
5	1202	94.3	238	5	Q17105	Q17105 aequeorea vi
6	1187	93.2	238	5	Q17106	Q17106 aequeorea vi
7	1082	84.9	238	5	Q8WTC6	Q8wtc6 aequeorea ma
8	1078	84.6	238	5	Q8WPP95	Q8wpc5 aequeorea ma
9	1074	84.3	238	5	Q8WTC4	Q8wpc4 aequeorea ma
10	1072	84.1	238	5	Q8WTD0	Q8wtd0 aequeorea ma
11	1071	84.1	238	5	Q8WTC8	Q8wtd8 aequeorea ma
12	1071	84.1	238	5	Q8WTC9	Q8wtd9 aequeorea ma
13	1069	83.9	238	5	Q8WTC7	Q8wtd7 aequeorea ma
14	1067	83.8	238	5	Q8WTC5	Q8wtd5 aequeorea ma
15	252.5	19.8	225	5	Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5	Q720W5	Q720w5 montastraea

..

```

Db      61 VTTFSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY      122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db      121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY      182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239
Db      181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 238

RESULT 2
Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-Optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,
RA Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 97.1%; Score 1237; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 8,7e-96;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVS GEGDATYKGLTLKFKICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVS GEGDATYKGLTLKFKICTTGKLPVWPPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 238

RESULT 3
Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324405; AAN86137.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.9%; Score 1235; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 1.3e-95;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVS GEGDATYKGLTLKFKICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVS GEGDATYKGLTLKFKICTTGKLPVWPPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHMKHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 238

RESULT 4
Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.9%; Score 1234; DB 2; Length 238;

```

```
Best Local Similarity 97.1%; Pred. No. 1.5e-95;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGSELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFCITTTGKLPVWPPTL 61
Db 1 MSKGSELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFCITTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121
Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMVLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMVLLEFVTAARITTHGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1BFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
DR NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.3%; Score 1202; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 7.5e-93;
Matches 223; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 VSKGSELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFCITTTGKLPVWPPTL 61
Db 1 MSKGSELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFCITTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121
Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMVLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMVLLEFVTAARITTHGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1BFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
DR NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 93.2%; Score 1187; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 1.3e-91;
Matches 221; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGSELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFCITTTGKLPVWPPTL 61
Db 1 MSKGSELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFCITTTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121
Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMVLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMVLLEFVTAARITTHGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; RAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR0009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
```


Query Match 83.8%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.9%; Pred.No. 1.6e-81;
Matches 195; Conservative 18; Mismatches 25; Indels 0; Gaps 0;
2 VSKGEETLTGCVDTLVELDGVNCHKFSVSGREGDATYKGLTLKFTCTTGKLPVPPPTL 61